

Testing mutual independence in high dimension via distance covariance

BY SHUN YAO, XIANYANG ZHANG, XIAOFENG SHAO *

Abstract

In this paper, we introduce a \mathcal{L}_2 type test for testing mutual independence and banded dependence structure for high dimensional data. The test is constructed based on the pairwise distance covariance and it accounts for the non-linear and non-monotone dependences among the data, which cannot be fully captured by the existing tests based on either Pearson correlation or rank correlation. Our test can be conveniently implemented in practice as the limiting null distribution of the test statistic is shown to be standard normal. It exhibits excellent finite sample performance in our simulation studies even when sample size is small albeit dimension is high, and is shown to successfully identify nonlinear dependence in empirical data analysis. On the theory side, asymptotic normality of our test statistic is shown under quite mild moment assumptions and with little restriction on the convergence rate of the dimension as a function of sample size. As a demonstration of good power properties for our distance covariance based test, we further show that an infeasible version of our test statistic has the rate optimality in the class of Gaussian distribution with equal correlation.

Keywords: *Banded dependence, Degenerate U-statistics, Distance correlation, High dimensionality, Hoeffding decomposition*

1 Introduction

In statistical multivariate analysis and machine learning research, a fundamental problem is to explore the relationships and dependence structure among subsets of variables. An

*Address correspondence to Xianyang Zhang (zhangxiany@tamu.edu), Assistant Professor, Department of Statistics, Texas A&M University. Shun Yao (shunyao2@illinois.edu) is Ph.D. candidate and Xiaofeng Shao (xshao@illinois.edu) is Professor at Department of Statistics, University of Illinois at Urbana-Champaign.

important dependence concept for a set of variables is mutual independence, which says that any two disjoint subsets of variables are independent from each other. Mutual independence can simplify the modeling and inference tasks of multivariate data considerably and certain models in multivariate analysis heavily depend on the mutual independence assumption. For example, in independent component analysis, it is often assumed that after a suitable linear transformation, the resulting set of variables are mutually independent. This paper is concerned with the testing of mutual independence of a p -dimensional random vector for a given random sample of size n . We are especially interested in the setting where $p > n$. This is motivated by the increasing statistical applications coming from biology, finance and neuroscience, among others, where the data dimension can be a lot larger than the sample size.

Given n independent observations $W_1, \dots, W_n \stackrel{D}{=} W$, where “ $\stackrel{D}{=}$ ” denotes equal in distribution and $W = (W^{(1)}, \dots, W^{(p)}) \sim \mathcal{F}$ with \mathcal{F} being a probability measure on the p dimensional Euclidean space, the goal is to test the mutual independence among the p components of W . That is to test the null hypothesis

$$H_0 : W^{(1)}, \dots, W^{(p)} \text{ are mutually independent}$$

versus

$$H_1 : \text{negation of } H_0.$$

To tackle this problem, one line of research focuses on the covariance matrices. Under the Gaussian assumption, testing H_0 is equivalent to testing that the covariance matrices are sphericity or identity after suitable scaling. When the dimension is fixed and smaller than the sample size, likelihood ratio tests [Anderson (1958)] and other multivariate tests [John (1971)] are widely used. In recent years, extensive works have emerged in the high dimensional context, where $p > n$, including Ledoit & Wolf (2002), Jiang (2004), Schott (2005), Srivastava (2005), Srivastava (2006), Chen et al. (2010), Fisher et al. (2010), Cai & Jiang (2011), Fisher (2012) among others. Existing tests can be generally categorized into two types: maximum type test [e.g. Cai & Jiang (2011)] and sum-of-squares (i.e. \mathcal{L}_2 type) test [e.g. Schott (2005), Chen et al. (2010)]. The former usually has an extreme distribution of type I and the latter has a normal limit under the null. For example, Cai & Jiang (2011) proved that their maximum Pearson correlation based statistic has an extreme distribution of type I under H_0 . Schott (2005), on the other hand, used the \mathcal{L}_2 type statistic with pairwise Pearson correlations, which attained a standard normal limiting null distribution.

It is well known that Pearson correlation cannot capture nonlinear dependence. To overcome this limitation, there have been some work based on rank correlation, which can

capture nonlinear albeit monotone dependence, and is also invariant to monotone transformation. For example, [Leung & Drton \(2015\)](#) proposed nonparametric tests based on sum of pairwise squared rank correlations in replacement of Pearson correlation in [Schott \(2005\)](#). They derived the standard normal limit under the regime where the ratio of sample size and dimension converges to a positive constant. [Han & Liu \(2014\)](#) considered a family of rank-based test statistics including the Spearman's rho and Kendall's tau correlation coefficients. Under the assumption that $\log p = o(n^{1/3})$, the limiting null distributions of their maximum type tests were shown to be an extreme value type I distribution.

Although rank correlation based test is distribution free and has some desirable finite sample properties, it has an intrinsic weakness, that is, it does not fully characterize dependence and it may have trivial power when the underlying dependence is non-monotonic. Furthermore, the maximum type statistics discussed above are known to converge to its theoretical limit at a very slow rate. This motivates us to use the distance covariance/correlation [[Székely et al. \(2007\)](#)] to quantify the dependence and build our test on the distance covariance. Distance correlation provides a natural extension of classical Pearson correlation and rank correlation in capturing arbitrary types of dependence. It measures the distance between the joint characteristic function of two random vectors of arbitrary dimensions and the product of their marginal characteristic functions in terms of weighted \mathcal{L}^2 norm. It has been shown in [Székely et al. \(2007\)](#) that distance correlation/covariance is zero if and only if the two random vectors are independent, thus it completely characterizes dependence.

The test statistic we consider is of the form

$$\hat{D}_n = \frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{\hat{S}},$$

where $dCov_n^2(W^{(l)}, W^{(m)})$ is the squared sample distance covariance between $W^{(l)}$ and $W^{(m)}$, and \hat{S} is a suitable studentizer defined later. Thus our test is of \mathcal{L}_2 type and it targets at non-sparse albeit weak pairwise dependence of any kind among the p components. It can be viewed as an extension of [Schott \(2005\)](#) and [Leung & Drton \(2015\)](#) by replacing Pearson correlation and rank correlation by distance covariance. Furthermore, our test statistic is later shown to be a degenerate U-statistic using the Hoeffding decomposition, which nevertheless admits a normal limit under both the null and (local) alternative hypothesis owing to the growing dimension.

Below we provide a brief summary of our contribution as well as some appealing features of our test. (1) Our test captures arbitrary type of pairwise dependence, which includes non-linear and non-monotone dependence that can be hardly detected by the existing tests for

mutual independence in the literature. (2) Our test does not involve any tuning parameters and uses standard normal critical value, so it can be conveniently implemented. (3) We develop the Hoeffding decomposition for the pairwise sample distance covariance which is an important step towards deriving the asymptotic distribution for the proposed test under some suitable assumptions. Our theoretical argument sheds some light on the behavior of U-statistics in the high dimensional settings and may have application to some other high dimensional inference problems. (4) Our test is shown to be rate optimal under the regime that p/n converges to a positive constant, when the data is from multivariate Gaussian with equal correlations. (5) We further extend the idea in testing mutual independence to test the banded dependence structure in high dimensional data, which is a natural follow-up testing procedure after the former test gets rejected.

It is worth noting that mutual (joint) independence implies mutual pairwise independence, but not vice versa. Thus our test, which actually tests for

$$H'_0 : W^{(1)}, \dots, W^{(p)} \text{ are mutually pairwise independent versus } H'_1 : \text{negation of } H'_0,$$

can fail to detect mutual dependence of more than two components. However, for a random vector with growing dimension, [Sun \(1998\)](#) showed that mutual pairwise and mutual independence are almost equivalent in a suitable sense, which provides a theoretical justification for our focus on mutual pairwise independence. Moreover, Theorem 11 of [Comon \(1994\)](#) showed that pairwise independence implies mutual independence in the context of independent component analysis. Our test is thus potentially useful in checking the independence among the estimated independent components. Section 5 provides some simulation evidence by comparing a test that aims to test mutual joint independence with ours, and indicates not much is lost by targeting mutual pairwise independence when p is large. In our testing context, pairwise dependence can be viewed as the main effect of mutual joint dependence, and dependence for triples and quadruples etc. can be regarded as high order interactions. Thus our test is consistent with the well-known statistical principle that we typically test for the presence of main effects before proceeding to the higher order interactions. In the absence of pairwise dependence, it is interesting that “almost all” triples are still approximately mutually independent; see the discussions on page 453 of [Sun \(1998\)](#).

The rest of the paper is organized as follows. Section 2 presents some preliminary results for distance covariance. Section 3 proposes the test statistic for testing mutual independence and studies its asymptotic properties under both the null and alternative. Section 4 describes an extension of the proposed test to testing the banded dependence. We provide several numerical comparisons in Section 5 and employ the proposed tests to analyze the prostate

cancer data set in Section 6. Section 7 concludes and sketches some future research directions. Appendix A contains all the technical details. All asymptotic results are stated under the framework that $\min(n, p) \rightarrow \infty$.

2 Preliminary: Distance covariance

The distance covariance between two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with finite first moments was first introduced by Székely et al. (2007). It is defined as the positive square root of

$$dCov^2(X, Y) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)|^2}{|t|_p^{1+p}|s|_q^{1+q}} dt ds,$$

where ϕ_X , ϕ_Y and $\phi_{X,Y}$ are the individual and joint characteristic functions of X and Y respectively, $|\cdot|_p$ and $|\cdot|_q$ are the Euclidean norms with the subscripts omitted later without ambiguity, $c_p = \pi^{(1+p)/2}/\Gamma((1+p)/2)$ is a constant and $\Gamma(\cdot)$ is the complete gamma function. Write $dCov^2(X) = dCov^2(X, X)$. The (squared) distance correlation is defined as a standardized version of (squared) distance covariance, i.e., $dCov^2(X, Y)/\sqrt{dCov^2(X)dCov^2(Y)}$ for $dCov^2(X)dCov^2(Y) > 0$, and it completely characterizes independence since it is zero if and only if X and Y are independent.

To obtain a suitable estimator for the squared distance covariance, we consider its alternative representation below. Let (X', Y') and (X'', Y'') be independent copies of (X, Y) . Further denote the double centered distance as $U(x, x') = |x - x'| - \mathbb{E}|x - X'| - \mathbb{E}|X - x'| + \mathbb{E}|X - X'|$ and $V(y, y') = |y - y'| - \mathbb{E}|y - Y'| - \mathbb{E}|Y - y'| + \mathbb{E}|Y - Y'|$, where x , x' , y and y' are dummy variables. According to Theorem 7 from Székely & Rizzo (2009), we have

$$\begin{aligned} \mathbb{E}U(X, X')V(Y, Y') &= \mathbb{E}|X - X'| |Y - Y'| - 2\mathbb{E}|X - X'| |Y - Y''| + \mathbb{E}|X - X'| \mathbb{E}|Y - Y'| \\ &= dCov^2(X, Y). \end{aligned}$$

Now given n random samples $Z_i = (X_i, Y_i) =^D (X, Y)$ for $i = 1, \dots, n$, we adopt the idea of \mathcal{U} -centering in Székely & Rizzo (2014) and Park et al. (2015) to construct an unbiased distance covariance estimator. Define $A = (A_{ij})_{i,j=1}^n$ and $B = (B_{ij})_{i,j=1}^n$, where $A_{ij} = |X_i - X_j|$ and $B_{ij} = |Y_i - Y_j|$. The \mathcal{U} -centered versions of A_{ij} and B_{ij} are defined respectively as

$$\begin{aligned} \tilde{A}_{ij} &= A_{ij} - \frac{1}{n-2} \sum_{l=1}^n A_{il} - \frac{1}{n-2} \sum_{k=1}^n A_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k,l=1}^n A_{kl}, \\ \tilde{B}_{ij} &= B_{ij} - \frac{1}{n-2} \sum_{l=1}^n B_{il} - \frac{1}{n-2} \sum_{k=1}^n B_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k,l=1}^n B_{kl}. \end{aligned}$$

An unbiased estimator of the (squared) distance covariance between X and Y is given by

$$dCov_n^2(X, Y) = \frac{1}{n(n-3)} \sum_{i \neq j} \tilde{A}_{ij} \tilde{B}_{ij}.$$

The following lemma shows that this estimator is a U-statistic and it is unbiased.

Lemma 2.1. *The sample distance covariance $dCov_n^2(X, Y)$ defined above is an unbiased estimator for $dCov^2(X, Y)$; Moreover, it is a fourth-order U-statistic which admits the form of*

$$dCov_n^2(X, Y) = \frac{1}{\binom{n}{4}} \sum_{i < j < k < l} h(Z_i, Z_j, Z_k, Z_l),$$

where

$$\begin{aligned} h(Z_i, Z_j, Z_k, Z_l) &= \frac{1}{4!} \sum_{(s,t,u,v)}^{(i,j,k,l)} (A_{st}B_{st} + A_{st}B_{uv} - 2A_{st}B_{su}) \\ &= \frac{1}{6} \sum_{s < t, u < v}^{(i,j,k,l)} (A_{st}B_{st} + A_{st}B_{uv}) - \frac{1}{12} \sum_{(s,t,u)}^{(i,j,k,l)} A_{st}B_{su} \end{aligned}$$

and the summation is over all permutations of the 4-tuples of indices (i, j, k, l) .

The variables $h(Z_i, Z_j, Z_k, Z_l)$ defined in Lemma 2.1 are not independent across $i < j < k < l$ which renders the derivation of asymptotic distribution a difficult task. Nevertheless, we shall adopt the classical Hoeffding decomposition, which provides a projection of U-statistic and separates out the dominant part that determines the asymptotic distribution of the U-statistic in the low dimensional setting. See [Serfling \(1980\)](#), [Lehmann \(1999\)](#) for more details. The proposition below states the Hoeffding decomposition for squared sample distance covariance.

Proposition 2.1. *Define $\nu^2 = \mathbb{E}U(X, X')^2 V(Y, Y')^2$ and $K(x, y) = \mathbb{E}U(x, X)V(y, Y)$. Assume that*

$$\mathbb{E}U(X, X'')^2 V(Y, Y')^2 = O(\nu^2), \tag{1}$$

$$dCov^2(X)dCov^2(Y) = O(\nu^2), \tag{2}$$

$$\text{var}(K(X, Y)) = o(n^{-1}\nu^2), \quad \text{var}(K(X, Y')) = o(\nu^2). \tag{3}$$

Then we have

$$dCov_n^2(X, Y) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} U(X_i, X_j)V(Y_i, Y_j) + \mathcal{R}_n, \tag{4}$$

where \mathcal{R}_n is the remainder term which is asymptotically negligible. When X and Y are independent, Conditions (1)-(3) hold automatically.

Remark 2.1. In the above proposition, $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are in arbitrary but fixed dimensions, whereas the dimension is allowed to grow in the following sections.

3 Mutual pairwise independence test

In the context of mutual independence testing, we denote n independent observations as $W_1, \dots, W_n =^D W$, where $W = (W^{(1)}, \dots, W^{(p)})$ and $W_i = (W_i^{(1)}, \dots, W_i^{(p)})$ for $i = 1, \dots, n$. We consider the following distance covariance based (infeasible) test statistic

$$D_n = \frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{S},$$

where S is a suitable studentizer to be defined later. Note that distance covariance has been used to test for independence between two random vectors; see [Székely & Rizzo \(2013a\)](#) and [Székely & Rizzo \(2013b\)](#).

To facilitate our derivation, we introduce some notation. Define the component-wise double centered distance $U_l(w^{(l)}, w'^{(l)}) = |w^{(l)} - w'^{(l)}| - \mathbb{E}|w^{(l)} - W'^{(l)}| - \mathbb{E}|W^{(l)} - w'^{(l)}| + \mathbb{E}|W^{(l)} - W'^{(l)}|$, where W' is an independent copy of W and $w, w' \in \mathbb{R}^p$ are dummy variables. Let $H(W_i, W_j) = \sum_{1 \leq l < m \leq p} U_l(W_i^{(l)}, W_j^{(l)}) U_m(W_i^{(m)}, W_j^{(m)})$. Notice that under the null $\mathbb{E}[H(W_i, W_j)] = \mathbb{E}[H(W_i, W_j)|W_i] = \mathbb{E}[H(W_i, W_j)|W_j] = 0$. Applying Proposition 2.1 to the pairwise distance covariance $dCov_n^2(W^{(l)}, W^{(m)})$, we obtain the following decomposition for our test statistic

$$D_n = \frac{1}{S \sqrt{\binom{n}{2}}} \sum_{1 \leq i < j \leq n} H(W_i, W_j) + \sum_{1 \leq l < m \leq p} \frac{\sqrt{\binom{n}{2}} \mathcal{R}_n^{(l,m)}}{S} = D_{n,1} + D_{n,2},$$

where $\mathcal{R}_n^{(l,m)}$ are the remainder terms for $1 \leq l < m \leq p$, and $D_{n,1}$ and $D_{n,2}$ are defined accordingly. To derive the asymptotic distribution of D_n , we use the results from Section 2 by replacing $U(X_i, X_j)V(Y_i, Y_j)$ with $H(W_i, W_j)$ in Proposition 2.1. It provides a neat and convenient way to control the remainder terms in the approximation.

First it is straightforward to show that

$$\text{var}(D_{n,1}) = \frac{1}{S^2 \binom{n}{2}} \sum_{1 \leq i < j \leq n} \sum_{1 \leq i' < j' \leq n} \mathbb{E} H(W_i, W_j) H(W_{i'}, W_{j'}) = \mathbb{E}[H(W, W')^2] / S^2.$$

Therefore, we shall choose $S^2 = \mathbb{E}[H(W, W')^2]$. Under the null, S^2 can be further simplified as

$$S^2 = \sum_{1 \leq l < m \leq p} dCov^2(W^{(l)}) dCov^2(W^{(m)}).$$

Using similar arguments from Appendix A.1.1, it can be shown that $D_{n,2}$ is asymptotically negligible under the null. Based on the above results, an unbiased estimator for S^2 under the null is

$$\hat{S}^2 = \sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)})dCov_n^2(W^{(m)}) \quad (5)$$

where $dCov_n^2(W^{(l)})$ is the unbiased estimator for $dCov^2(W^{(l)})$ as defined in Section 2.

Therefore, we consider the following feasible test statistic

$$\hat{D}_n = \frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{\hat{S}}.$$

In Section 3.1, we establish the asymptotic normality for our test statistic \hat{D}_n under the null, which leads to the following decision rule for our testing procedure

$$\phi_{n,\alpha}(W_1, \dots, W_n) := \begin{cases} 1 & \text{if } \hat{D}_n > z_\alpha \\ 0 & \text{if } \hat{D}_n \leq z_\alpha, \end{cases}$$

where z_α is the $100(1 - \alpha)\%$ quantile of standard normal. We reject the null hypothesis H_0 if $\phi_{n,\alpha} = 1$, not reject otherwise.

3.1 Asymptotic analysis under the null of mutual independence

To derive the asymptotic distribution for the proposed test statistic \hat{D}_n under the null, we introduce the following assumptions

Assumption A1. As $n \rightarrow \infty$ and $p \rightarrow \infty$,

$$\frac{\sum_{l=1}^p \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4 + n^{-1} \sum_{l=1}^p \text{var}(W^{(l)})^2}{\{\sum_{l=1}^p dCov^2(W^{(l)})\}^2} \rightarrow 0,$$

where $\mu^{(l)} = \mathbb{E}[W^{(l)}]$.

Notice that the first term in assumption A1 can also be rewritten as $\sum_{l=1}^p \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4 / S^2 \rightarrow 0$ as we showed in the Appendix A.2.2 that $\frac{1}{2}[\sum_{l=1}^p dCov^2(W^{(l)})]^2$ is the leading term in the variance, that is, $S^2 = \frac{1}{2}[\sum_{l=1}^p dCov^2(W^{(l)})]^2 \{1 + o(1)\}$ under the null. Therefore, we assume that the sum of the p components' first centered absolute moments to the fourth power grows at a slower rate than S^2 , and the sum of the p components' squared variance grows at most $o(n)$ faster than S^2 . This is in fact a very mild assumption. For example,

when the element-wise second moments and the distance variances are all lower and upper bounded uniformly, as in the standard multivariate Gaussian case, the above assumption is trivially satisfied. Note that there is no explicit relationship between p and n in the above assumption, and they are allowed to grow independently.

To further appreciate Assumption A1, we mention Assumptions B1-B2 below, which involve more explicit convergence rate and admit more direct interpretation. It is easy to see that Assumptions B1-B2 imply Assumption A1, which suffices for our asymptotic analysis.

Assumption B1.

$$\liminf_{p \rightarrow \infty} \frac{1}{p} \sum_{l=1}^p dCov^2(W^{(l)}) > 0.$$

Assumption B2.

$$\limsup_{p \rightarrow \infty} \frac{1}{p} \sum_{l=1}^p var(W^{(l)})^2 < \infty.$$

Assumption B1 is a mild assumption on the joint distribution of W . The inequality sets a lower bound on the average distance variance of the p components of W . Notice that $dCov^2(X) = 0$ if and only if X is a constant. Therefore, it basically assumes that at least a non-negligible portion of the components of W are not constants. Assumption B2 is also fairly mild, which only requires that the average of squared variance across the p components of W is finite. It is weaker than the assumption that the variance of each component of W is uniformly bounded.

Proposition A.1 in the Appendix provides us a useful tool to derive the asymptotic distribution for our test statistic. We can therefore use the central limit theorem for sum of martingale difference sequences [Hall (1984)] to derive the asymptotic distribution for the infeasible test statistic D_n , as stated below.

Theorem 3.1. *Under the null hypothesis H_0 and Assumption A1, we have*

$$D_n := \frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{S} \xrightarrow{d} N(0, 1),$$

as $p \rightarrow \infty$ and $n \rightarrow \infty$.

We obtained the asymptotic normality for the infeasible statistic D_n without imposing any explicit or implicit constraints on the convergence rates of the dimension p and sample size n , and both can grow to infinity freely. In our feasible test statistic, we replace S^2 by its unbiased estimator \hat{S}^2 as defined in equation (5). We show the ratio consistency of the above variance estimator in the next theorem.

Theorem 3.2. *Under the null hypothesis H_0 and Assumption A1, we have*

$$\frac{\hat{S}^2}{S^2} \xrightarrow{p} 1 \quad \text{as } p \text{ and } n \rightarrow \infty.$$

Comparing to Theorem 3.1, we do not impose any additional assumptions in obtaining the ratio-consistency. Then we can combine Theorem 3.1 and Theorem 3.2, and derive the asymptotic normality of \hat{D}_n by applying Slutsky's theorem.

Corollary 3.1. *Under the null hypothesis H_0 and Assumption A1, we have*

$$\hat{D}_n := \frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{\hat{S}} \xrightarrow{d} N(0, 1)$$

as $p \rightarrow \infty$ and $n \rightarrow \infty$.

It is worth highlighting that our test is developed in a model free setting. No parametric/nonparametric model was assumed and only weak distributional assumptions are required. The second moment assumptions seem necessary given the fact that our test is built on sample distance covariance. It is indeed possible to relax the moment assumptions further by considering the so-called ranked distance covariance, i.e., replacing sample distance covariance by the sample ranked distance covariance, which is obtained by applying distance covariance to the ranks for any two components, say the ranks based on $(W_1^{(l)}, \dots, W_n^{(l)})$ and $(W_1^{(m)}, \dots, W_n^{(m)})$, respectively. Additionally, it is possible to combine the idea of aggregation with other tests developed for independence of two univariate random variables (see e.g., Heller et al. (2013), Heller et al. (2016)) and form a test for mutual pairwise independence. These extensions are beyond the scope of this paper and are left for future research.

3.2 Asymptotic analysis under the alternatives

Now we focus on the local alternatives where some pairs among the p components are dependent, i.e., $dCov^2(W^{(l)}, W^{(m)}) > 0$ for some $l, m \in \{1, \dots, p\}$. Let

$$D'_n = S^{-1} \sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} \{dCov_n^2(W^{(l)}, W^{(m)}) - dCov^2(W^{(l)}, W^{(m)})\}, \quad (6)$$

where $S^2 = \mathbb{E}[H(W, W')^2]$.

By the Hoeffding decomposition, we have $D'_n := D'_{n,1} + \sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} \mathcal{R}_n^{(l,m)} / S$ where

$$D'_{n,1} := \frac{1}{S \sqrt{\binom{n}{2}}} \sum_{1 \leq i < j \leq n} \{H(W_i, W_j) - \mathbb{E}H(W_i, W_j)\}.$$

and the contribution from the remainder term $\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} \mathcal{R}_n^{(l,m)} / S$ is asymptotically negligible under the assumptions,

$$\mathbb{E} [T(W, W', W'')]^2 = O(S^2), \quad (7)$$

$$\mathbb{E} \left[\sum_{1 \leq l < m \leq p} U_l(W^{(l)}, W'^{(l)}) U_m(W''^{(m)}, W'''^{(m)}) \right]^2 = O(S^2), \quad (8)$$

$$\text{var} (\mathbb{E}_W(T(W, W', W'))) = o(n^{-1} S^2), \quad \text{var} (\mathbb{E}_W(T(W, W', W''))) = o(S^2), \quad (9)$$

where we define $T(W, W', W'') = \sum_{1 \leq l < m \leq p} U_l(W^{(l)}, W'^{(l)}) U_m(W''^{(m)}, W'^{(m)})$ and \mathbb{E}_W denotes the expectation with respect to W .

Conditions (7)–(9) are obtained from (1)–(3) and they characterize the local alternative we discuss here in an abstract way. Notice that under the null of mutual independence, these conditions are automatically satisfied and $\text{var} (\mathbb{E}_W(T(W, W', W'))) = 0$ in Condition (9), which makes our test statistic a degenerate U-statistic under the null. For the local alternative, we also focus on the degenerate case in the sense that we require the alternative not too far away from the null. Therefore, these conditions guarantee that our test statistic is still degenerate when some pairs among the p components are dependent. In the case that $\text{var} (\mathbb{E}_W(T(W, W', W'))) \neq 0$ and the test statistic is non-degenerate, we can regard it as the fixed alternative; its asymptotic distribution can be derived similarly under suitable assumptions.

Furthermore, we can rewrite $D'_{n,1}$ under Condition (9) using the double centered version of $H(W, W')$ as

$$D'_{n,1} = \frac{1}{S \sqrt{\binom{n}{2}}} \sum_{1 \leq i < j \leq n} \tilde{H}(W_i, W_j) + o_p(1),$$

where $\tilde{H}(W, W') = H(W, W') - \mathbb{E}[H(W, W')|W] - \mathbb{E}[H(W, W')|W'] + \mathbb{E}[H(W, W')]$. Similar to the arguments under the null in Section 3.1 and Proposition A.2–A.3 in the Appendix, we define the following quantities

$$\begin{aligned} \tilde{\mathcal{V}}_1 &= \mathbb{E}[\tilde{H}(W, W')^2 \tilde{H}(W, W'')^2], \\ \tilde{\mathcal{V}}_2 &= \mathbb{E}[\tilde{H}(W, W') \tilde{H}(W, W'') \tilde{H}(W''', W') \tilde{H}(W''', W'')], \\ \tilde{\mathcal{V}}_3 &= \mathbb{E}[\tilde{H}(W, W')^4], \\ \tilde{S}^2 &= \mathbb{E}[\tilde{H}(W, W')^2]. \end{aligned}$$

The following theorem establishes the asymptotic normality for D'_n based on similar arguments under the null with H replaced by \tilde{H} .

Theorem 3.3. *Under the Conditions (7)–(9) and also*

$$\frac{\tilde{\mathcal{V}}_1}{n\tilde{S}^4} \rightarrow 0, \quad \frac{\tilde{\mathcal{V}}_2}{\tilde{S}^4} \rightarrow 0, \quad \frac{\tilde{\mathcal{V}}_3}{n^2\tilde{S}^4} \rightarrow 0, \quad (10)$$

we have $D'_n \rightarrow^d N(0, 1)$.

Using Theorem 3.3, we can readily show that the power function of the test statistic \hat{D}_n is approximately

$$\Phi \left(-z_\alpha + \sqrt{\binom{n}{2}} \sum_{1 \leq l < m \leq p} dCov^2(W^{(l)}, W^{(m)})/S \right)$$

where $\Phi(\cdot)$ and z_α are the distribution function and $100(1 - \alpha)\%$ quantile of standard normal respectively. If $\sum_{1 \leq l < m \leq p} dCov^2(W^{(l)}, W^{(m)}) = O(S)$, for example, we have a non-trivial power function approximately $\Phi(-z_\alpha + O(n))$.

Remark 3.1. *Condition (10) stated in Theorem 3.3 is quite abstract due to the fact that we are working under a very abstract class of local alternative characterized by Conditions (7)–(9). It seems difficult to make these conditions more explicit without extra or more specific assumptions under the local alternative. Actually it is possible to conduct a more detailed analysis under a more specific local alternative class. For example, consider the following local alternative class, $\mathcal{L}_k := \{(W^{(1)}, \dots, W^{(p)}) \mid \text{the indices } \{1, 2, \dots, p\} \text{ can be divided into two subsets } \mathcal{C}_k \text{ and } \mathcal{C}_k^c \text{ with } |\mathcal{C}_k| = k; \text{ for } i \in \mathcal{C}_k, W^{(i)} \text{ is dependent with at least one component with index in } \mathcal{C}_k; (W^{(i)})_{i \in \mathcal{C}_k^c} \text{ are mutually independent; } (W^{(i)})_{i \in \mathcal{C}_k} \text{ and } (W^{(i)})_{i \in \mathcal{C}_k^c} \text{ are independent}\}$. This particular local alternative class basically assumes only a small portion of the components are dependent relative to the dimension p . It can be shown that it belongs to the abstract local alternative characterized by Conditions (7)–(9) above and the asymptotic normality can be derived with some moment assumptions. The details are lengthy so are not presented here but are available upon request.*

3.3 Rate optimality under Gaussian equicorrelation

Usually when the signal or dependence is too weak, it may be very difficult to distinguish between the null and the alternative hypothesis. In this subsection, we study the boundary for the testable, non-testable region and also show that our test is optimal under the testable region.

We focus on the case where $W = (W^{(1)}, \dots, W^{(p)})$ follows a p -variate Gaussian distribution. Without loss of generality, we assume each of the marginals is standard Gaussian

with unit variance. Then our null hypothesis is equivalent to $\Sigma - I_p = 0$. We introduce the following alternative class $\mathcal{N}_p(\|\Sigma - I_p\|_F \geq c)$ which was also discussed in [Cai & Ma \(2013\)](#), [Leung & Drton \(2015\)](#),

$$\mathcal{N}_p(\|\Sigma - I_p\|_F \geq c) := \{W = (W^{(1)}, \dots, W^{(p)}) | W \sim N_p(\mu, \Sigma), \|\Sigma - I_p\|_F \geq c\}$$

where $N_p(\mu, \Sigma)$ denotes a p -variate Gaussian distribution with mean μ and covariance matrix Σ , $\|\cdot\|_F$ is the matrix Froebenius norm and I_p is the $p \times p$ identity matrix. Here $\|\Sigma - I_p\|_F$ quantifies the signal/dependence strength. From Theorem 1 in [Cai & Ma \(2013\)](#) we know that if c is sufficiently small, then no α -level test can distinguish between the null and alternative with desired power under the regime that p/n is bounded. Therefore any test that can achieve arbitrary large power for large enough c under this regime will be rate-optimal. The following results show the rate optimality for our proposed test.

Consider the equicorrelation alternative class $\mathcal{N}_p^{equi}(\|\Sigma - I_p\|_F \geq c)$, which is a sub-class of \mathcal{N}_p such that all the pairwise correlations equal to a common value denoted as ρ . Let $\Theta = (dCov(W^{(l)}, W^{(m)}))_{1 \leq l < m \leq p}$ be the $\binom{p}{2}$ -vector of all the pairwise distance covariance. It is easy to see that $\mathcal{N}_p^{equi}(\|\Sigma - I_p\|_F \geq c)$ is equivalent to $\mathcal{N}_p^{equi}(|\Theta| \geq \tilde{c})$ for some \tilde{c} . Here we use the fact that for standard Gaussian variables with correlation ρ , we have

$$dCov^2(W^{(l)}, W^{(m)}) = \frac{4}{\pi} [\rho \arcsin \rho + \sqrt{1 - \rho^2} - \rho \arcsin(\rho/2) - \sqrt{4 - \rho^2} + 1] := f(\rho). \quad (11)$$

In view of the proof of Theorem 7 in [Székely et al. \(2007\)](#), we have $c_1 \rho^2 p^2 \leq |\Theta|^2 = \sum_{l < m} dCov^2(W^{(l)}, W^{(m)}) \leq c_2 \rho^2 p^2$ for some positive constants c_1 and c_2 . With a slight abuse of notation, we shall use $\phi_{n,\alpha}$ to denote the decision rule based on our infeasible test statistic D_n in this subsection.

Theorem 3.4. *For any $0 < \alpha < \beta < 1$, as $p/n \rightarrow \lambda \in (0, \infty)$, there exists a constant $\tilde{c} = \tilde{c}(\alpha, \beta, \lambda) > 0$ such that*

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{N}_p^{equi}(|\Theta| \geq \tilde{c})} \mathbb{E}[\phi_{n,\alpha}] > \beta.$$

We conjecture that the same result presented in Theorem 3.4 also holds for the feasible test statistic \hat{D}_n , but it seems very involved to derive a probabilistic bound for $\hat{S}^2/S^2 - 1$, which is required in the proof. Nevertheless, the above result suggests that our distance covariance based test has potentially good power properties in the special case of Gaussian distributions, as shared by rank correlation based test of [Leung & Drton \(2015\)](#). See Section 5 for numerical evidence.

4 Testing for banded dependence structure

We propose a test statistic in this section to test for the banded dependence structure. Usually when the null hypothesis of mutual independence is rejected, it is of interest to test for some specific dependence structure afterwards or independently. For example, when the p components have a natural ordering, which arises in time series analysis, testing for m -dependence is of particular interest. Moreover, in the high dimensional covariance matrix estimation literature, banded covariance structure attracts a lot of attention; see [Wu & Pourahmadi \(2003\)](#), [Bickel & Levina \(2008\)](#), [Wagaman & Levina \(2009\)](#), [Shao & Zhou \(2014\)](#) among others. [Qiu & Chen \(2012\)](#) built a test for banded covariance matrices and also presented an approach to estimating the corresponding bandwidth; [Cai & Jiang \(2011\)](#), [Han & Liu \(2014\)](#) used Pearson correlation and rank correlation respectively for testing banded linear and monotone dependence. In contrast, our proposed test for banded structure targets any kind of dependence using distance covariance as analogous to the mutual independence test in Section 3. Accordingly, we consider the following null hypothesis for the banded dependence structure:

$$H_{0,h} : W^{(l)} \text{ and } W^{(m)} \text{ are independent for all } |l - m| \geq h.$$

Define

$$H_h(W_i, W_j) = \sum_{\substack{1 \leq l < m \leq p \\ |l - m| \geq h}} U_l(W_i^{(l)}, W_j^{(l)}) U_m(W_i^{(m)}, W_j^{(m)}).$$

Then the (infeasible) distance covariance based statistic for testing $H_{0,h}$ is

$$D_{n,h} = S_h^{-1} \sum_{\substack{1 \leq l < m \leq p \\ |l - m| \geq h}} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)}),$$

where $S_h^2 = \mathbb{E}[H_h(W, W')^2]$. The variance estimator we consider here is

$$\hat{S}_h^2 = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \left(\sum_{\substack{1 \leq l < m \leq p \\ |l - m| \geq h}} \tilde{A}_{ij}(l) \tilde{A}_{ij}(m) \right)^2.$$

Similarly define \tilde{S}_h^2 , $\mathcal{V}_{j,h}$ and $\tilde{\mathcal{V}}_{j,h}$ as the h -lag analogues of \tilde{S}^2 , \mathcal{V}_j and $\tilde{\mathcal{V}}_j$ for $1 \leq j \leq 3$ from Section 3. Following similar arguments in the proofs of Theorem 3.1 and Theorem 3.3, we have the following theorem for testing banded dependence structure.

Theorem 4.1. Define $T_h(W, W', W'') = \sum_{\substack{1 \leq l < m \leq p \\ |l-m| \geq h}} U_l(W^{(l)}, W''^{(l)}) U_m(W^{(m)}, W'^{(m)})$. Then under the assumptions that

$$\mathbb{E} [T_h(W, W', W'')]^2 = O(S_h^2), \quad (12)$$

$$\mathbb{E} \left[\sum_{\substack{1 \leq l < m \leq p \\ |l-m| \geq h}} U_l(W^{(l)}, W'^{(l)}) U_m(W''^{(m)}, W'''^{(m)}) \right]^2 = O(S_h^2), \quad (13)$$

$$\text{var}(\mathbb{E}_W(T_h(W, W', W''))) = o(n^{-1} S_h^2), \quad \text{var}(\mathbb{E}_W(T_h(W, W', W''))) = o(S_h^2) \quad (14)$$

and also

$$\frac{\tilde{\mathcal{V}}_{1,h}}{n\tilde{S}_h^4} \rightarrow 0, \quad \frac{\tilde{\mathcal{V}}_{2,h}}{\tilde{S}_h^4} \rightarrow 0, \quad \frac{\tilde{\mathcal{V}}_{3,h}}{n^2\tilde{S}_h^4} \rightarrow 0, \quad (15)$$

we have

$$D'_{n,h} := S_h^{-1} \sum_{\substack{1 \leq l < m \leq p \\ |l-m| \geq h}} \sqrt{\binom{n}{2}} [dCov_n^2(W^{(l)}, W^{(m)}) - dCov^2(W^{(l)}, W^{(m)})] \rightarrow^d N(0, 1).$$

Furthermore, under the null hypothesis of banded dependence, $\mathbb{E}_W(T_h(W, W', W')) = 0$ and condition (14) is satisfied automatically; \tilde{S}_h^2 , $\tilde{\mathcal{V}}_{j,h}$ reduce to S_h^2 and $\mathcal{V}_{j,h}$ for $1 \leq j \leq 3$. We have $D_{n,h} \rightarrow^d N(0, 1)$.

Similar to the discussion in Section 3.2, the theorem is presented under an abstract local alternative class characterized by (12) - (14). These conditions can be further studied under a more specific definition of the local alternative class, which we did not pursue here in this paper.

5 Simulation

In this section, we conduct Monte Carlo simulations to assess the finite sample performance of the mutual independence test in Section 5.1, the banded dependence test in Section 5.2 and present a comparison between pairwise independence test with a joint independence test in Section 5.3; we also compare our proposed methods (dCov, hereafter) with the following existing tests in the literature. Schott (2005) proposed a \mathcal{L}_2 type statistic using pairwise Pearson correlation (SC, hereafter); Leung & Drton (2015) studied the \mathcal{L}_2 type statistics using either Kendall's tau (LD_τ) or Spearman's rho (LD_ρ); Along a different line Cai & Jiang (2011) used the \mathcal{L}_∞ type statistic of Pearson Correlation to test the structure of

covariance matrices (CJ test, hereafter); Han & Liu (2014) developed the \mathcal{L}_∞ type statistics using either Kendall’s tau (HL_τ) or Spearman’s rho (HL_ρ). In Section 5.2, adaptations of the CJ and HL_ρ tests to testing the banded dependence structure are also carried out to compare with the proposed test. Section 5.3 provides more details about dHSIC, a testing procedure proposed by Pfister et al. (2016) to target joint independence.

5.1 Testing for Mutual independence

In this subsection, we evaluate the size and power of the proposed mutual independence test for both Gaussian and non-Gaussian distributions. The size and power (rejection probabilities) reported below are based on 5000 Monte Carlo simulations at the nominal level $\alpha = 0.05$. We choose sample size $n = \{60, 100\}$ and the dimension $p = \{50, 100, 200, 400, 800\}$.

Example 5.1. *The data $W = (W_1, \dots, W_p) \in \mathbb{R}^p$ are generated as follows with each component independent from others*

- **i)** *The data are generated from a standard Gaussian distribution with $W \sim N_p(0, I_p)$;*
- **ii)** *The data are generated from a Gaussian copula family with $W = Z^{1/3}$ and $Z \sim N_p(0, I_p)$;*
- **iii)** *The data are generated from a Gaussian copula family with $W = Z^3$ and $Z \sim N_p(0, I_p)$;*
- **iv)** *The components $\{W_j\}_{j=1}^p$ are i.i.d. from the student-t distribution with degrees of freedom three.*

The sizes for all the tests are summarized in Table 1. The performance of the proposed test is very comparable to those from LD_τ and LD_ρ . SC’s test performs reasonably well in cases i) and ii), especially when the underlying data is Gaussian. However, it has slightly upward size inflation in case iv) and exhibits severe size distortion in case iii). The \mathcal{L}_∞ type statistics HL_τ and HL_ρ turn out to be conservative for all the scenarios; CJ’s test has an unpleasantly high rejection rate in cases iii) and iv) due to the violation of Gaussian assumption. In addition, Figure 5.1 shows the histogram of the test statistics from 5000 Monte Carlo simulation of case i) as well as the kernel density estimate using the Gaussian kernel. Comparing with the red dashed line (density of standard normal), we observe that the null distribution of the test statistics is in general very close to standard normal for all the combinations of (n, p) being considered.

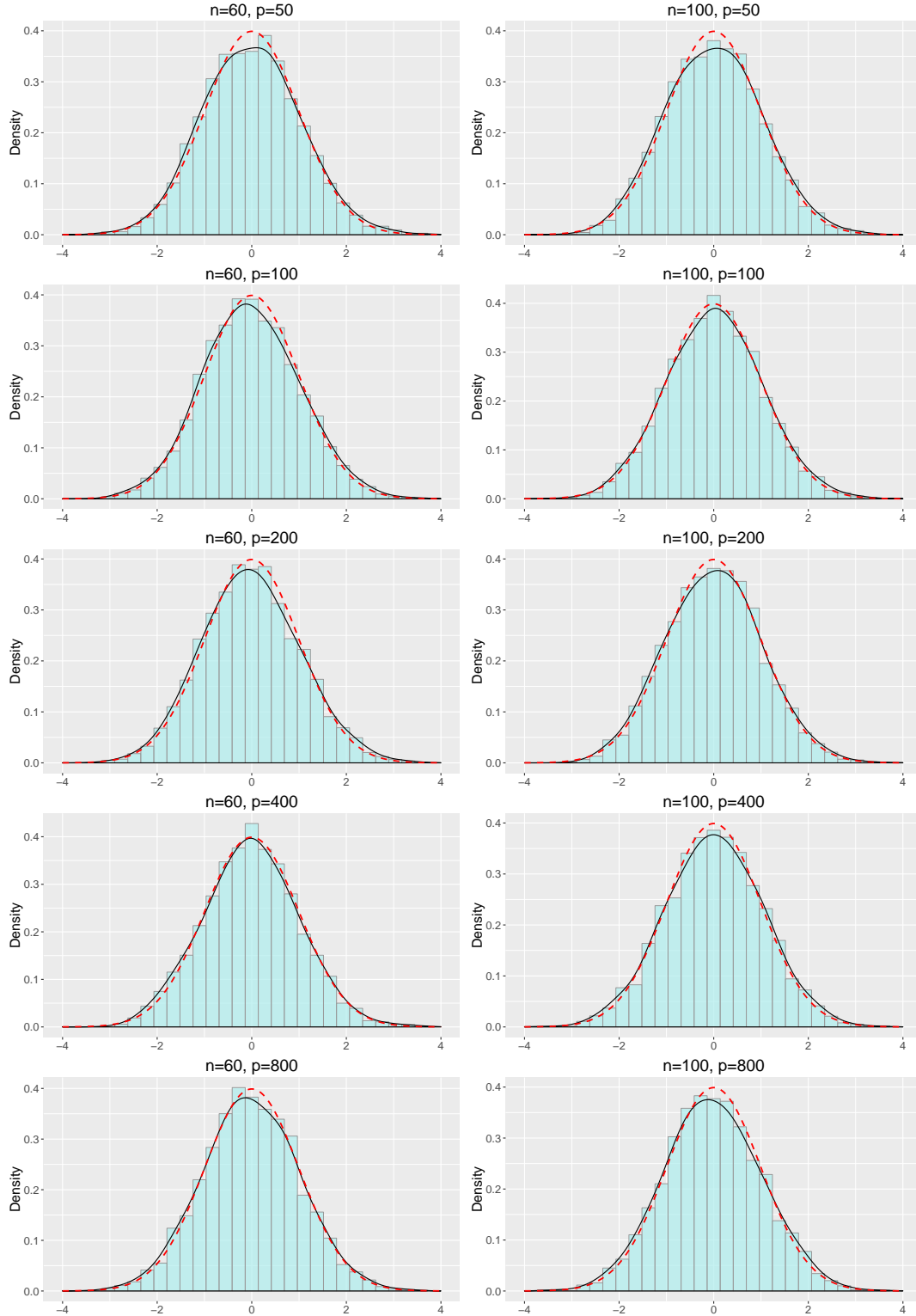


Figure 5.1: The histogram and kernel density estimate for the null distribution of the test statistics for Example 5.1. The red dashed line is the density of the standard normal.

Example 5.2. The data $W = (W_1, \dots, W_p) \in \mathbb{R}^p$ are generated from multivariate normal distribution with the following three covariance matrices $\Sigma = (\sigma_{ij}(\rho))_{i,j=1}^p$ for $\rho = 0.25$.

- **AR(1) structure:** $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho^{|i-j|}$ for all $i, j \in \{1, \dots, d\}$;
- **Band structure:** $\sigma_{ii} = 1$ for $i = 1, \dots, d$; $\sigma_{ij} = \rho$ if $0 < |i - j| < 3$ and $\sigma_{ij} = 0$ if $|i - j| \geq 3$;
- **Block structure:** Define $\Sigma_{\text{block}} = (\sigma_{ij}^*)$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$ if $i \neq j$ for all $i, j \in \{1, \dots, 5\}$. The covariance matrix is given by the following Kronecker product $\Sigma = \mathbf{I}_{\lfloor p/5 \rfloor} \otimes \Sigma_{\text{block}}$.

Table 2 reports the power from Example 5.2. It shows that the \mathcal{L}_2 type tests perform equally well with power one for most of the cases, while the maximum type tests endure severe power loss when sample size is small or dimension is high. The reason lies in the fact that the alternatives we consider here are dense and therefore favor the \mathcal{L}_2 type tests, whereas the \mathcal{L}_∞ type tests target sparse alternative instead and do not work very well in this case.

Example 5.3. Let ω be generated from a standard Gaussian distribution with $\omega \sim N_{p/5}(0, I_{p/5})$. The dependence structure is constructed through the non-linear functions such that $W = (g_1(\omega), g_2(\omega), g_3(\omega), g_4(\omega), g_5(\omega)) \in \mathbb{R}^p$, where $g_1(x) = x$, $g_2(x) = \sin(2\pi x)$, $g_3(x) = \cos(2\pi x)$, $g_4(x) = \sin(4\pi x)$ and $g_5(x) = \cos(4\pi x)$ and $g_i(\omega)$ means applying the function g_i to each component of ω .

Example 5.4. Let ω be generated from a standard Gaussian distribution with $\omega \sim N_{p/2}(0, I_{p/2})$. The dependence structure is constructed through the non-linear functions such that $W = (g_1(\omega), g_2(\omega)) \in \mathbb{R}^p$, where $g_1(x) = x$ and $g_2(x) = \log(x^2)$ and $g_i(\omega)$ means applying the function g_i to each component of ω .

Example 5.5. Let ω be generated from univariate standard normal distribution. The dependence structure is constructed through the non-linear functions such that $W = (\sin(\pi\omega), \sin(2\pi\omega), \dots, \sin(p\pi\omega))$.

Examples 5.3, 5.4 and 5.5 are designed for the non-linear and non-monotone dependence, in which case our dCov-based test demonstrates the highest power among all the competing methods as seen from Table 3. In particular, the power from the proposed test increases as sample size and dimension increase. However, other three \mathcal{L}_2 type tests only exhibit power in Example 5.5 and the powers diminish substantially and even down to nominal level

in other cases. On the other hand, for the \mathcal{L}_∞ type tests, only HL_τ has some power in detecting the non-monotone dependence; the other two maximum type tests maintain the power around nominal level α . These examples clearly demonstrate the advantage of the distance covariance based test in identifying the non-linear and non-monotone dependence among the data.

5.2 Testing for banded dependence structure

In this subsection, we conduct additional simulation to evaluate the performance of the proposed test extended to testing for the banded dependence structure. The simulation setting is the same as in Section 5.1.

Example 5.6. *Consider the following banded dependence structures*

- i) *The data is generated from multivariate normal distribution with banded covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^p$, where $\sigma_{ii} = 1$ for $i = 1, \dots, d$, $\sigma_{ij} = 0.3$ if $0 < |i - j| < 5$ and $\sigma_{ij} = 0$ if $|i - j| \geq 5$;*
- ii) *The data is generated as $W = Z^3$, where Z is generated from i);*
- iii) *The data is generated as $W = Z^{1/3}$, where Z is generated from i).*

Table 4 shows the result from Example 5.6. The true bandwidth is 4 in this example, we choose $h = 5$ and $h = 10$ in the tests. It can be seen from the table that dCov-based banded dependence structure test has slight size inflation when $n = 60$, which subsides as sample size grows. In contrast, HL_τ test is a little bit conservative in some scenarios. CJ test is more conservative in case i, iii and shows strong size distortion when the distribution is too far from Gaussian in case ii. It appears that there is no big difference between using $h = 5$ and $h = 10$ for all of the three tests. Likewise, we provide the histogram of the test statistics from 5000 Monte Carlo simulation and also the kernel density estimate using the Gaussian kernel with the comparison of standard normal density as the red dashed line in Figure 5.2 for the three cases in this example where $n = 100$, $p = 800$ and $h = 10$. It is shown that the normal approximation is quite close to the null distribution of the proposed test statistic in all the three cases. The plots for $h = 5$ are almost identical to those for $h = 10$ and therefore omitted.

Example 5.7. *Consider the following cases*

- i) *The data is generated from multivariate normal distribution with banded covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^p$, where $\sigma_{ii} = 1$ for $i = 1, \dots, d$, $\sigma_{ij} = 0.1$ if $0 < |i - j| \leq 20$ and $\sigma_{ij} = 0$ if $|i - j| > 20$;*

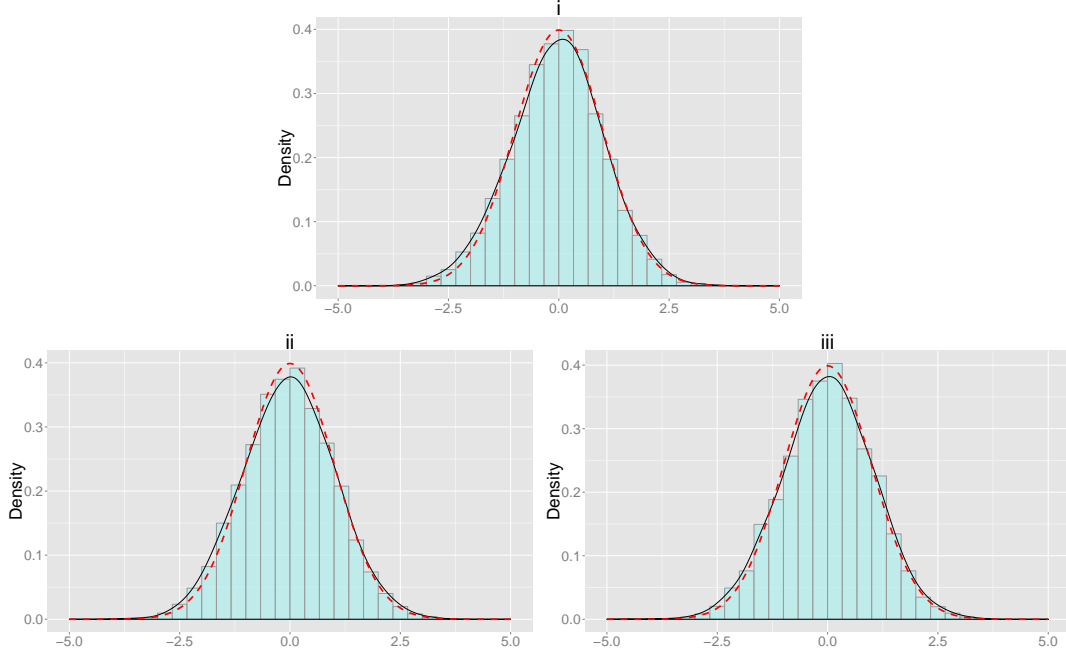


Figure 5.2: The histogram and kernel density estimate for the null distribution of the test statistics for Example 5.6. The red dashed line is the density of the standard normal.

- ii) The data is generate as $W = Z^{1/3}$, where Z is generated from i).
- iii) The data is generate the same way as Example 5.3 in Section 5.1.

Table 5 collects the results from Example 5.7. We choose $h = 5$ and $h = 10$ in all the tests whereas the true bandwidths in case i and ii are both 20; in case iii, there is no banded dependence structure. We observe that the power for the proposed test is consistently higher than other methods and the power increases as sample size and dimension increase. HL_τ test suffers from significant power reduction in all cases. Moreover, CJ test is the worst among these three tests with power less than the nominal level in most of the scenarios. This example demonstrates that our proposed banded dependence test has very good power performance.

5.3 Pairwise independence versus joint independence

As mentioned in the introduction, our test mainly focuses on the presence of the “main effects” of joint (mutual) dependence and tests for the sub-null H'_0 . In comparison, Pfister et al. (2016) proposed a testing procedure (dHSIC, hereafter) based on the idea of Hilbert-Schmidt independence criterion and their kernel-based test targets at the joint (mutual) independence. Note that the theory for dHSIC is restricted to the fixed dimensional case and

its validity in the high dimensional case is unknown. Here we compare our proposed method with dHSIC under different scenarios, and see whether there is any size/power discrepancy between the two methods.

Since dHSIC test requires that $n \geq 2p$, we choose three combinations $n = 60, p = 18$; $n = 100, p = 36$ and $n = 200, p = 72$. We compare the two tests for some of the examples chosen from Section 5.1, namely Example 5.1, 5.2, 5.4 and 5.5. Besides, we also consider an interesting example as follows, where W is pairwise independent but not jointly independent.

Example 5.8. Consider the tuple of three random variables $\mathbf{Z} = (Z_1, Z_2, Z_3)$, where Z_1, Z_2 are independent Bernoulli random variables with success probability $1/2$, $Z_3 = \mathbf{1}_{(Z_1=Z_2)}$ and $\mathbf{1}_{(\cdot)}$ is the indicator function. Our data consists of $p/3$ i.i.d copies of \mathbf{Z} , that is, $W = (\mathbf{Z}_1, \dots, \mathbf{Z}_{p/3})$.

The size and power (rejection probabilities) are reported based on 5000 Monte Carlo simulations at the nominal level $\alpha = 0.05$. Here the dHSIC is implemented as a permutation test using Gaussian kernel, and the bandwidth parameter is chosen as the median of the sample distances. Note that distance covariance has been shown as a special case of kernel distance in [Sejdinovic et al. \(2013\)](#).

Table 6 summarizes the rejection rates for our distance covariance based test and dHSIC test. For $(n, p) = (200, 72)$, dHSIC delivers zero rejection rates in some cases. A careful look at the source code indicates that the critical values are set to be ∞ in these cases so the resulting rejection rates are zero. For smaller (n, p) , the performance of dHSIC seems more reasonable and we shall comment on that below. When the data are jointly independent as in Example 5.1, both tests have quite accurate rejection rates around the nominal level 5%, which suggests that normal approximation works quite well for our test even when $p = 18$ and $n = 60$. For linear dependent and non-linear dependent data in Example 5.2 and Example 5.4-5.5 respectively, dCov demonstrates consistently high power against the null; surprisingly, dHSIC almost has no power for linear dependent data and has very little power in Example 5.4. For Example 5.8, the data is pairwise independent but not jointly independent, thus our test cannot detect any dependence beyond the pairwise dependence and has rejection rate around the nominal level, which is consistent with our expectation; dHSIC has a reasonable rejection rate when dimension is small relative to the sample size, but endures severe power loss when the dimension is high. The fact that the power for dHSIC is so low when $(n, p) = (200, 72)$ is somewhat expected, since most of triples in Example 5.8 are mutually independent (as mentioned in [Sun \(1998\)](#)) and thus the data with dimension $p = 72$ are less mutually dependent than that when $p = 36$ and 18. These findings suggest that

the incapability of the dHSIC to handle high dimensional data, and there may be intrinsic difficulty to capture all kinds of high order dependence beyond pairwise dependence when the dimension is high.

6 Data Illustration

In this section, we employ the proposed methods to analyze the prostate cancer data set and report the results. The original prostate cancer data was analyzed by [Adam et al. \(2002\)](#) to study the protein profiling technologies that can simultaneously resolve and analyze multiple proteins in early detection of prostate cancer. Surface enhanced laser desorption/ionization mass spectrometry protein profiles of patients' blood serum samples are recorded. These profiles contains the intensity values for a large amount of time-of-flight values. The time-of-flight is related to the mass over charge ratio m/z of the constituent proteins in the blood. There are 157 healthy patients and 167 prostate cancer patients with 48,538 m/z -sites in total.

This data set has been analyzed by several statisticians for various purposes. Following previous researchers, the m/z -sites below 2000 are ignored due to the possible chemical artifacts occurrence under that level. [Tibshirani et al. \(2005\)](#) averaged the intensity values in consecutive blocks of 20, which gives a total of 2181 dimensions per serum sample. [Levina et al. \(2008\)](#), [Qiu & Chen \(2012\)](#) further averaged the data of [Tibshirani et al. \(2005\)](#) in consecutive blocks of 10, resulting in a total of 218 dimensions. We follow this approach and consider the observation $\mathbf{w}_i = (w_{i,1}, \dots, w_{i,218})$ with intensity profile of length 218 for patient i to test the mutual independence and the banded dependence structure if the former hypothesis is rejected.

We conduct the analysis for two groups separately: the healthy group (157 samples), prostate cancer group (167 samples). The tests for mutual independence are both rejected for these two groups with p-values 0. Since there is a natural ordering for these 218 dimensions (m/z -sites), we further carry out the banded dependence structure test with given bandwidth h from 50 to 217. The corresponding values of the test statistics are plotted in [Figure 6](#). We also employ the proposed methods to the mixed group data (157 healthy patients together with 167 prostate cancer patients), but the results are not informative and therefore omitted. Some previous studies also used the standardized data and we found no significant differences between using the original data and the standardized data in our tests for this particular prostate cancer data set.

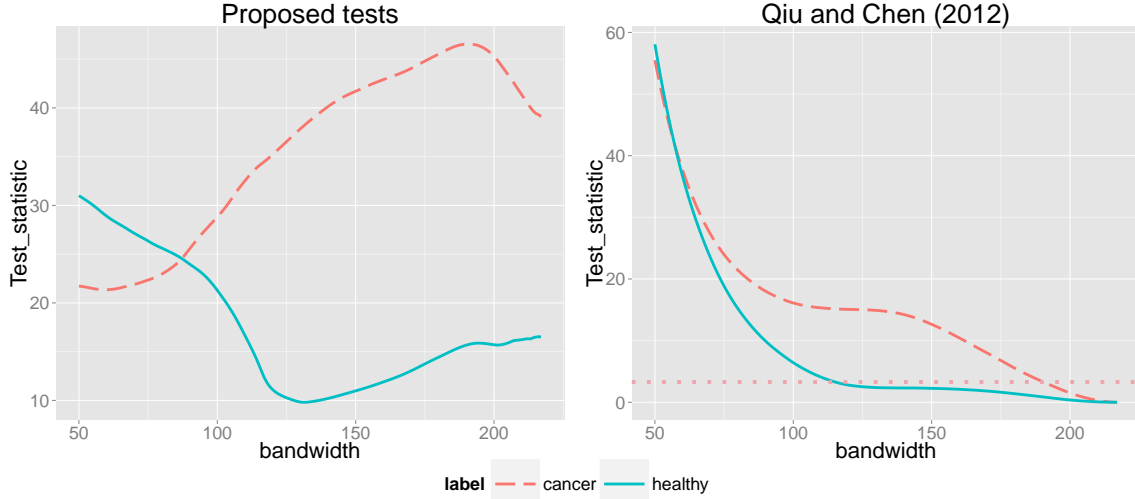


Figure 6.1: Values of test statistics for healthy and prostate cancer patients of the proposed test (left panel) and [Qiu & Chen \(2012\)](#) test (right panel).

The test results suggest that the dependence structure is not banded for both the patient group and healthy group. The shape of the curve from healthy group in the left panel of Figure 6 indicates that the overall dependence is decreasing steeply first and then increasing gradually as the bandwidth increases; moreover, the test statistics are the smallest for $125 \leq h \leq 150$, which hints at that the dependence is relatively weak for those bandwidths. The curve from prostate cancer group, however, demonstrates completely different pattern. It increases substantially from $h = 70$ to $h = 185$ and then decrease afterwards, which suggests strong non-linear dependence. The sharp contrast between healthy group and cancer group signifies significant differences in the dependence structure for prostate cancer and non-cancer people.

[Bickel & Levina \(2008\)](#), [Qiu & Chen \(2012\)](#) (the test statistic values are shown in the right panel of Figure 6) used covariance matrix based method and concluded that the healthy group's covariance matrix is likely to be banded with bandwidth 144 and 121 respectively and may not be banded at all for the prostate cancer group. Our method implies that the dependence structure is not banded for both groups and the non-linear dependence is especially strong between bandwidth 90 and 185 for the cancer group.

7 Conclusion

In the present paper, we proposed a mutual independence test using pairwise squared distance covariance and further extended the test to testing the banded dependence structure. Asymptotic distributions of the test statistics were studied under the null and local alternatives using tools related to U-statistics. We view our new test as a useful addition to the family of mutual independence tests, for example, [Schott \(2005\)](#), [Cai & Jiang \(2011\)](#), [Han & Liu \(2014\)](#), [Leung & Drton \(2015\)](#) among others, as none of the existing tests can capture non-monotonic dependence. Our numerical results demonstrate the merit of the proposed test in identifying the non-linear and non-monotonic dependence in the data compared with Pearson correlation and rank correlation based counterparts, which only focus on linear dependence and monotone dependence respectively.

As mentioned early, sum of squares/ \mathcal{L}^2 type statistic naturally targets at non-sparse but weak alternatives. It would be interesting to consider the \mathcal{L}_∞ /maximum type statistic using the distance covariance in the future to capture sparse and strong dependence. The mild size distortion for our test at small sample size may be alleviated by using permutation-based critical values. However, permutation based test becomes quite expensive in high dimension, and it will be interesting to develop more accurate approximation of our null distribution with manageable/scalable computational cost. Furthermore, we can use distance correlation based test in testing mutual independence or consider a more general multivariate dependence measure instead of pairwise dependence measure to capture the dependence of any three or more subsets of p components, which is certainly more challenging and is left for future work.

8 Acknowledgement

Zhang acknowledges partial support from NSF grant DMS-1607320. Shao would like to acknowledge partial financial support from National Science Foundation grants DMS-1407037 and DMS-1607489. We are grateful to Dr. Niklas Pfister for pointing us to the “dHSIC” package in CRAN.

A Technical Appendix

A.1 Hoeffding decomposition

Define that $h_c(z_1, \dots, z_c) = \mathbb{E}h(z_1, \dots, z_c, Z_{c+1}, \dots, Z_4)$, where $Z_i = (X_i, Y_i) \stackrel{D}{=} (X, Y)$ for $c = 1, 2, 3, 4$. Let $z = (x, y)$, $z' = (x', y')$, $z'' = (x'', y'')$ and $z''' = (x''', y''')$. Let (X', Y') , (X'', Y'') and (X''', Y''') be independent copies of (X, Y) . Direct calculation yields that

$$\begin{aligned}
 h_1(z) &= \frac{1}{2} \left\{ \mathbb{E}|x - X|(|Y' - Y''| + |y - Y| - |y - Y'| - |Y - Y'|) \right. \\
 &\quad \left. + \mathbb{E}|X - X'|(|y - Y''| + |Y - Y'| - |y - Y| - |Y' - Y''|) \right\} \\
 &= \frac{1}{2} \left\{ \mathbb{E}|x - X|[V(Y', Y'') + V(y, Y) - V(y, Y') - V(Y, Y'')] \right. \\
 &\quad \left. + \mathbb{E}|X - X'|[V(y, Y'') + V(Y, Y') - V(y, Y) - V(Y', Y'')] \right\} \\
 &= \frac{1}{2} \left\{ \mathbb{E}(|x - X| - |X - X'|)[V(y, Y) - V(Y, Y'')] + \mathbb{E}|X - X'|V(Y, Y') \right\} \\
 &= \frac{1}{2} \left\{ \mathbb{E}U(x, X)V(y, Y) + dCov^2(X, Y) \right\}.
 \end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
 h_2(z, z') &= \frac{1}{6} \left\{ U(x, x')V(y, y') + dCov^2(X, Y) \right. \\
 &\quad \left. + \mathbb{E}U(x, X)(2V(y, Y) - V(y', Y)) \right. \\
 &\quad \left. + \mathbb{E}U(x', X)(2V(y', Y) - V(y, Y)) \right\},
 \end{aligned}$$

and

$$\begin{aligned}
 h_3(z, z', z'') &= \frac{1}{12} \left\{ (2U(x, x') - U(x', x'') - U(x, x''))V(y, y') \right. \\
 &\quad + (2U(x, x'') - U(x, x') - U(x', x''))V(y, y'') \\
 &\quad + (2U(x', x'') - U(x, x') - U(x, x''))V(y', y'') \\
 &\quad + \mathbb{E}(2U(x, X) - U(x', X) - U(x'', X))V(y, Y) \\
 &\quad + \mathbb{E}(2U(x', X) - U(x, X) - U(x'', X))V(y', Y) \\
 &\quad \left. + \mathbb{E}(2U(x'', X) - U(x, X) - U(x', X))V(y'', Y) \right\},
 \end{aligned}$$

and

$$\begin{aligned}
& h_4(z, z', z'', z''') \\
&= \frac{1}{12} \left\{ (2U(x, x') + 2U(x'', x''') - U(x, x'') - U(x, x''') - U(x', x'') - U(x', x'''))(V(y, y') + V(y'', y''')) \right. \\
&\quad + (2U(x, x'') + 2U(x', x''') - U(x, x') - U(x, x''') - U(x'', x') - U(x'', x'''))(V(y, y'') + V(y', y''')) \\
&\quad \left. + (2U(x, x''') + 2U(x'', x') - U(x, x'') - U(x, x') - U(x''', x'') - U(x''', x'))(V(y, y''') + V(y', y'')) \right\}.
\end{aligned}$$

A.1.1 Proof of Proposition 2.1

Proof. Under the null of mutual independence between X and Y , $dCov^2(X, Y) = 0$. It can be easily seen that $h_1(z) = 0$. And h_2 and h_3 can be simplified as,

$$h_2(z, z') = \frac{1}{6}U(x, x')V(y, y'),$$

and

$$\begin{aligned}
h_3(z, z', z'') &= \frac{1}{12} \left\{ (2U(x, x') - U(x', x'') - U(x, x''))V(y, y') \right. \\
&\quad + (2U(x, x'') - U(x, x') - U(x', x''))V(y, y'') \\
&\quad \left. + (2U(x', x'') - U(x, x') - U(x, x''))V(y', y'') \right\}.
\end{aligned}$$

We deduce that

$$\text{var}(h_2(Z, Z')) = \frac{1}{36}\mathbb{E}U(X, X')^2V(Y, Y')^2 := \nu^2,$$

and

$$\begin{aligned}
\text{var}(h_3(Z, Z', Z'')) &= \frac{3}{144}\text{var}\{(2U(X, X') - U(X', X'') - U(X, X''))V(Y, Y')\} \\
&= \frac{1}{24} \left[2\mathbb{E}U(X, X')^2V(Y, Y')^2 + \mathbb{E}U(X, X'')^2V(Y, Y')^2 \right] \\
&= \frac{1}{8}\mathbb{E}U(X, X')^2V(Y, Y')^2 \\
&= O(\nu^2),
\end{aligned}$$

and also

$$\begin{aligned}
\text{var}(h_4(Z, Z', Z'', Z''')) &= \frac{6}{144} \mathbb{E}V(Y, Y')^2 [U(X, X'') + U(X', X''') + U(X', X'') \\
&\quad + U(X, X''') - 2U(X, X') - 2U(X'', X''')]^2 \\
&= \frac{1}{6} \{ \mathbb{E}V(Y, Y')^2 U(X, X'')^2 + \mathbb{E}U(X, X')^2 \mathbb{E}V(Y, Y')^2 \\
&\quad + \mathbb{E}U(X, X')^2 V(Y, Y')^2 \} \\
&= \frac{1}{2} \mathbb{E}U(X, X')^2 V(Y, Y')^2 \\
&= O(\nu^2).
\end{aligned}$$

The sample distance covariance can be decomposed as in (4) under the null. The readers are referred to [Serfling \(1980\)](#) for more details.

Under the local alternative, we assume that

$$\text{var}(K(X, Y)) = o(n^{-1}\nu^2), \quad \text{var}(K(X, Y')) = o(\nu^2).$$

This condition implies that

$$\text{var}(h_1(Z)) = o(n^{-1}\nu^2), \quad \text{var}(h_2(Z, Z')) = \nu^2(1 + o(1)).$$

Moreover, we have

$$\begin{aligned}
\text{var}(h_3(Z, Z', Z'')) &\leq C \left\{ \nu^2 + \mathbb{E}U(X, X'')^2 V(Y, Y')^2 + \mathbb{E}U(X, X'')U(X', X'')V(Y, Y')^2 \right\} \\
&\leq C \left\{ \nu^2 + \mathbb{E}U(X, X'')^2 V(Y, Y')^2 \right\},
\end{aligned}$$

and

$$\text{var}(h_4(Z, Z', Z'', Z''')) \leq C' \left\{ \nu^2 + \mathbb{E}U(X, X'')^2 V(Y, Y')^2 + \mathbb{E}U(X, X')^2 \mathbb{E}V(Y, Y')^2 \right\},$$

where C and C' are some constants which are independent of n and p . Therefore, the same decomposition can be derived under assumptions (1)-(3). \square

A.2 Proofs of the main results

A.2.1 Proof of Lemma 2.1

Proof. Denote $\mathbf{1} \in \mathbb{R}^n$ as the vector of all ones, $(n)_k = n!/(n-k)!$, I_k^n is the collections of k -tuples of indices from $\{1, 2, \dots, n\}$ such that each index occurs only once. By Lemma 1 of

Park et al. (2015), it can be shown that

$$\begin{aligned}
dCov_n^2(X, Y) &= \frac{1}{n(n-3)} \left(\text{tr}(AB) + \frac{\mathbf{1}^T A \mathbf{1} \mathbf{1}^T B \mathbf{1}}{(n-1)(n-2)} - \frac{2\mathbf{1}^T A B \mathbf{1}}{(n-2)} \right) \\
&= (n)_4^{-1} \sum_{(i,j,k,l) \in I_4^n} (A_{ij}B_{ij} + A_{ij}B_{kl} - 2A_{ij}B_{ik}) \\
&= \frac{1}{\binom{n}{4}} \sum_{i < j < k < l} h(Z_i, Z_j, Z_k, Z_l)
\end{aligned}$$

where

$$\begin{aligned}
h(Z_i, Z_j, Z_k, Z_l) &= \frac{1}{4!} \sum_{(s,t,u,v)}^{(i,j,k,l)} (A_{st}B_{st} + A_{st}B_{uv} - 2A_{st}B_{su}) \\
&= \frac{1}{6} \sum_{s < t, u < v}^{(i,j,k,l)} (A_{st}B_{st} + A_{st}B_{uv}) - \frac{1}{12} \sum_{(s,t,u)}^{(i,j,k,l)} A_{st}B_{su}
\end{aligned}$$

with $Z_i = (X_i, Y_i)$, and the last summation is over all permutations of the 4-tuples of indices (i, j, k, l) . It is straightforward to verify that

$$\begin{aligned}
\mathbb{E} \left[\sum_{(i,j) \in I_2^n} A_{ij}B_{ij} \right] &= \mathbb{E}[\text{tr}(AB)] = (n)_2 \cdot \mathbb{E}|X - X'| |Y - Y'|, \\
\mathbb{E} \left[\sum_{(i,j,q,r) \in I_4^n} A_{ij}B_{qr} \right] &= \mathbb{E}[\mathbf{1}^T A \mathbf{1} \mathbf{1}^T B \mathbf{1} - 4\mathbf{1}^T A B \mathbf{1} + 2\text{tr}(AB)] = (n)_4 \cdot \mathbb{E}|X - X'| \mathbb{E}|Y - Y'|, \\
\mathbb{E} \left[\sum_{(i,j,r) \in I_3^n} A_{ij}B_{ir} \right] &= \mathbb{E}[\mathbf{1}^T A B \mathbf{1} - \text{tr}(AB)] = (n)_3 \cdot \mathbb{E}|X - X'| |Y - Y''|.
\end{aligned}$$

Therefore, $dCov_n^2(X, Y)$ is unbiased and it is a fourth-order U-statistic. \square

A.2.2 Proof of Theorem 3.1

Define the following quantities,

$$\begin{aligned}
\mathcal{V}_1 &= \mathbb{E}[H(W, W')^2 H(W, W'')^2], \\
\mathcal{V}_2 &= \mathbb{E}[H(W, W') H(W, W'') H(W''', W') H(W''', W'')], \\
\mathcal{V}_3 &= \mathbb{E}[H(W, W')^4].
\end{aligned}$$

We first present the following three propositions.

Proposition A.1. Define $M_r := \sum_{j=2}^r \sum_{i=1}^{j-1} H(W_i, W_j)$. Then M_r is a martingale relative to the natural filtration with respect to $\{W_i\}_{i=1}^r$.

Proof. Define the natural filtration $\mathcal{F}_j = \sigma(W_1, W_2, \dots, W_j)$. Notice that under the null

$$\mathbb{E}[H(W_i, W_j)] = \mathbb{E}[H(W_i, W_j)|W_i] = \mathbb{E}[H(W_i, W_j)|W_j] = 0.$$

It follows that $M_r \in \mathcal{F}_r$ and $\mathbb{E}(M_r) = 0$. For any $s \geq r$,

$$\begin{aligned} \mathbb{E}(M_s|\mathcal{F}_r) &= \sum_{j=2}^r \sum_{i=1}^{j-1} H(W_i, W_j) + \mathbb{E} \left[\sum_{j=r+1}^s \sum_{i=1}^{j-1} H(W_i, W_j) \middle| \mathcal{F}_r \right] \\ &= M_r + \sum_{j=r+1}^s \sum_{i=1}^r \mathbb{E} [H(W_i, W_j)|\mathcal{F}_r] + \sum_{j=r+2}^s \sum_{i=r+1}^{j-1} \mathbb{E} [H(W_i, W_j)] \\ &= M_r + \sum_{j=r+1}^s \sum_{i=1}^r \mathbb{E} \left[\sum_{1 \leq l < m \leq p} U_l(W_i^{(l)}, W_j^{(l)}) U_m(W_i^{(m)}, W_j^{(m)}) | W_i \right] \\ &= M_r. \end{aligned}$$

Therefore, M_r is a zero mean martingale sequence. \square

Proposition A.2. Define $\mathcal{W}_j = \sum_{i=1}^{j-1} H(W_i, W_j)$ and the natural filtration \mathcal{F}_j with respect to W_j . Then under the assumption that

$$\frac{\mathcal{V}_1}{nS^4} \rightarrow 0, \quad \frac{\mathcal{V}_2}{S^4} \rightarrow 0, \quad (16)$$

we have

$$B_n^{-2} \sum_{j=2}^n \mathbb{E}(\mathcal{W}_j^2 | \mathcal{F}_{j-1}) \rightarrow^p 1, \quad (17)$$

where $B_n^2 = n(n-1)S^2/2$.

Proof of Proposition A.2. Notice that

$$\begin{aligned} \sum_{j=2}^n \mathbb{E}[\mathcal{W}_j^2] &= \sum_{j=2}^n \mathbb{E} \left[\sum_{i, i'=1}^{j-1} \sum_{1 \leq l < m \leq p} U_l(W_i^{(l)}, W_j^{(l)}) U_m(W_i^{(m)}, W_j^{(m)}) \right. \\ &\quad \cdot \left. \sum_{1 \leq l' < m' \leq p} U_{l'}(W_{i'}^{(l')}, W_j^{(l')}) U_{m'}(W_{i'}^{(m')}, W_j^{(m')}) \right] \\ &= \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E} H(W_i, W_j)^2 \\ &= \frac{n(n-1)}{2} S^2 = B_n^2. \end{aligned}$$

Define $L_j(W_i, W_k) = \mathbb{E}[H(W_i, W_j)H(W_k, W_j)|\mathcal{F}_{j-1}]$ for $i, k < j$, and note that

$$\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}] = \mathbb{E}\left[\sum_{i=1}^{j-1} \sum_{k=1}^{j-1} H(W_i, W_j)H(W_k, W_j)|\mathcal{F}_{j-1}\right] = \sum_{i=1}^{j-1} \sum_{k=1}^{j-1} L_j(W_i, W_k).$$

If $i \leq k$ and $i' \leq k'$ then

$$\begin{aligned} & \mathbb{E}[L_j(W_i, W_k)L_{j'}(W_{i'}, W_{k'})] \\ &= \mathbb{E}H(W, W')^2 H(W, W'')^2 && \text{if } i = k = i' = k', \\ &= \mathbb{E}[H(W, W')H(W, W'')H(W''', W')H(W''', W'')] && \text{if } i = i' \neq k = k', \text{ or } i = k' \neq k = i', \\ &= [\mathbb{E}H(W, W')^2]^2 && \text{if } i = k \neq i' = k', \\ &= 0 && \text{otherwise,} \end{aligned}$$

and also

$$\begin{aligned} & \mathbb{E}[L_j(W_i, W_k)]\mathbb{E}[L_{j'}(W_{i'}, W_{k'})] \\ &= \mathbb{E}H(W_i, W_j)H(W_k, W_j)\mathbb{E}H(W_{i'}, W_{j'})H(W_{k'}, W_{j'}) \\ &= [\mathbb{E}H(W, W')^2]^2 && \text{if } i = k, \ i' = k', \\ &= 0 && \text{otherwise.} \end{aligned}$$

Therefore,

$$\begin{aligned} \text{var} \left(\sum_{j=2}^n \mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}] \right) &= \sum_{j,j'=2}^n \sum_{i,k=1}^{j-1} \sum_{i',k'=1}^{j'-1} \text{cov}(L_j(W_i, W_k), L_{j'}(W_{i'}, W_{k'})) \\ &= \sum_{j=j'} [(j-1)\mathcal{V}_1 + 2(j-1)(j-2)\mathcal{V}_2 - (j-1)S^4] \\ &\quad + 2 \sum_{2 \leq j \leq j' \leq n} [(j-1)\mathcal{V}_1 + 2(j-1)(j-2)\mathcal{V}_2 - (j-1)S^4]. \end{aligned}$$

Under the assumption (16), we have

$$\frac{4}{n^2(n-1)^2 S^4} \text{var} \left(\sum_{j=2}^n \mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}] \right) \rightarrow 0.$$

Therefore (17) holds. □

Proposition A.3. Define $\mathcal{W}_j = \sum_{i=1}^{j-1} H(W_i, W_j)$ and the natural filtration \mathcal{F}_j with respect to W_j . Under the assumption

$$\frac{\mathcal{V}_1}{nS^4} \rightarrow 0, \quad \frac{\mathcal{V}_3}{n^2 S^4} \rightarrow 0, \tag{18}$$

we have

$$\sum_{j=2}^n B_n^{-2} \mathbb{E} (\mathcal{W}_j^2 \mathbf{I}(|\mathcal{W}_j| > \epsilon B_n) | \mathcal{F}_{j-1}) \rightarrow^p 0, \quad (19)$$

where $B_n = n(n-1)\mathcal{S}^2/2$.

Proof of Proposition A.3. Notice that

$$\sum_{j=2}^n B_n^{-2} \mathbb{E} (\mathcal{W}_j^2 \mathbf{1}(|\mathcal{W}_j| > \epsilon B_n) | \mathcal{F}_{j-1}) \leq \sum_{j=2}^n B_n^{-2} (\epsilon B_n)^{-s} \mathbb{E} (|\mathcal{W}_j|^{2+s} | \mathcal{F}_{j-1})$$

for some $s > 0$. It suffices to show that for $s = 2$

$$\sum_{j=2}^n B_n^{-4} \mathbb{E} (\mathcal{W}_j^4 | \mathcal{F}_{j-1}) \rightarrow^p 0.$$

To this end, we show that

$$\sum_{j=2}^n B_n^{-4} \mathbb{E} (\mathcal{W}_j^4) \rightarrow^p 0. \quad (20)$$

Some algebra yields that

$$\begin{aligned} \sum_{j=2}^n \mathbb{E}[\mathcal{W}_j^4] &= \sum_{j=2}^n \sum_{i_1, i_2, i_3, i_4=1}^{j-1} \mathbb{E} H(W_{i_1}, W_j) H(W_{i_2}, W_j) H(W_{i_3}, W_j) H(W_{i_4}, W_j) \\ &= \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E}[H(W_i, W_j)^4] + 3 \sum_{j=2}^n \sum_{i \neq i'}^{j-1} \mathbb{E}[H(W_i, W_j)^2 H(W_{i'}, W_j)^2] \\ &= \frac{n(n-1)}{2} \mathcal{V}_3 + O(n^3 \mathcal{V}_1). \end{aligned}$$

Therefore, under (18), (20) holds. \square

We present the following lemma which is useful in the proof of Theorem 3.1.

Lemma A.1. *Let $a(x) = \max\{|\mathbb{E}[|X - X'|] - 2\mathbb{E}[|x - X'|]|, \mathbb{E}[|X - X'|]\}$. Then we have $|U(x, x')| \leq \max\{a(x), a(x')\}$.*

Proof of Lemma A.1. By the triangle inequality, we have $|\mathbb{E}[|X - x'|] - |x - x'|| \leq \mathbb{E}[|x - X'|]$ for $x, x' \in \mathbb{R}$. Thus $|U(x, x')| \leq \max\{|\mathbb{E}[|X - X'|] - 2\mathbb{E}[|x - X'|]|, \mathbb{E}[|X - X'|]\} = a(x)$. Switching x and x' , we get $|U(x, x')| \leq a(x')$. The conclusion thus follows. \square

Proof of Theorem 3.1. We show that Assumption A1 implies both (16) and (18) under the null, i.e., $\frac{\mathcal{V}_1}{nS^4} \rightarrow 0$, $\frac{\mathcal{V}_2}{S^4} \rightarrow 0$ and $\frac{\mathcal{V}_3}{n^2S^4} \rightarrow 0$. We write $a \lesssim b$ if a is less or equal to b up to a

multiplicative constant. By Lemma A.1 and the fact that $\mathbb{E}[a(X)] \lesssim E[|X - \mathbb{E}[X]|]$, we have

$$\begin{aligned} \frac{\sum_{l=1}^p dCov^4(W^{(l)})}{[\sum_{l=1}^p dCov^2(W^{(l)})]^2} &= \frac{\sum_{l=1}^p \{\mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2]\}^2}{[\sum_{l=1}^p dCov^2(W^{(l)})]^2} \\ &\leq \frac{\sum_{l=1}^p \{\mathbb{E}[a(W^{(l)})]\}^4}{[\sum_{l=1}^p dCov^2(W^{(l)})]^2} \\ &\lesssim \frac{\sum_{l=1}^p \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4}{[\sum_{l=1}^p dCov^2(W^{(l)})]^2}. \end{aligned}$$

By Assumption A1, $\sum_{l=1}^p dCov^4(W^{(l)}) = o([\sum_{l=1}^p dCov^2(W^{(l)})]^2)$. Therefore, we have

$$\begin{aligned} 2S^2 &= \sum_{l \neq m} dCov^2(W^{(l)}) dCov^2(W^{(m)}) \\ &= \left\{ \sum_{l=1}^p dCov^2(W^{(l)}) \right\}^2 - \sum_{l=1}^p dCov^4(W^{(l)}) \\ &= \left\{ \sum_{l=1}^p dCov^2(W^{(l)}) \right\}^2 \cdot \{1 + o(1)\}. \end{aligned}$$

Again using Lemma A.1 and the fact that $\mathbb{E}[a(X)^2] \lesssim \text{var}(X)$, we have

$$\begin{aligned} \mathcal{V}_1 &= \mathbb{E}[H(W, W')^2 H(W, W'')^2] \\ &= \sum_{l < m} \sum_{l' < m'} \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_m(W^{(m)}, W'^{(m)})^2 U_{l'}(W^{(l')}, W''^{(l')})^2 U_{m'}(W^{(m')}, W''^{(m')})^2] \\ &\lesssim \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2] \right\}^4 + \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_l(W^{(l)}, W''^{(l)})^2] \right\}^2 \\ &\quad + \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_l(W^{(l)}, W''^{(l)})^2] \right\} \cdot \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2] \right\}^2 \\ &\lesssim \left\{ \sum_l dCov^2(W^{(l)}) \right\}^4 + \left\{ \sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^2 \text{var}(W^{(l)}) \right\}^2. \end{aligned}$$

Together with Assumption A1, we can show that

$$\frac{\mathcal{V}_1}{nS^4} \lesssim \left\{ \sum_l dCov^2(W^{(l)}) \right\}^4 \cdot \frac{1}{nS^4} + \left\{ \sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^2 \text{var}(W^{(l)}) \right\}^2 \cdot \frac{1}{nS^4} \rightarrow 0,$$

where we have used the Cauchy-Schwarz inequality to show $\{\sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^2 \text{var}(W^{(l)})\}^2 \leq$

$\sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4 \sum_l \text{var}(W^{(l)})^2$. Similarly, we have

$$\begin{aligned}
\mathcal{V}_2 &= \mathbb{E}[H(W, W')H(W, W'')H(W''', W')H(W''', W'')] \\
&= \sum_{1 \leq l < m \leq p} \mathbb{E}[U_l(W^{(l)}, W'^{(l)})U_l(W^{(l)}, W''^{(l)})U_l(W'''^{(l)}, W'^{(l)})U_l(W'''^{(l)}, W''^{(l)}) \\
&\quad \cdot U_m(W^{(m)}, W'^{(m)})U_m(W^{(m)}, W''^{(m)})U_m(W'''^{(m)}, W'^{(m)})U_m(W'''^{(m)}, W''^{(m)})] \\
&\leq \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})U_l(W^{(l)}, W''^{(l)})U_l(W'''^{(l)}, W'^{(l)})U_l(W'''^{(l)}, W''^{(l)})] \right\}^2 \\
&\lesssim \left\{ \sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4 \right\}^2,
\end{aligned}$$

which implies that

$$\frac{\mathcal{V}_2}{S^4} \leq \left\{ \frac{1}{S^2} \sum_l \{\mathbb{E}[|W^{(l)} - \mu^{(l)}|]\}^4 \right\}^2 \rightarrow 0.$$

Lastly, we have

$$\begin{aligned}
\mathcal{V}_3 &= \mathbb{E}[H(W, W')^4] \lesssim \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^4] \right\}^2 + \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2] \right\}^4 \\
&\quad + \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^4] \right\} \cdot \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2] \right\}^2 \\
&\quad + \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2] \right\} \cdot \left\{ \sum_l \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^3] \right\}^2 \\
&\lesssim \left\{ \sum_l \text{var}(W^{(l)})^2 \right\}^2 + \left\{ \sum_l dCov^2(W^{(l)}) \right\}^4,
\end{aligned}$$

Hence,

$$\frac{\mathcal{V}_3}{n^2 S^4} \lesssim \left\{ \frac{1}{n S^2} \sum_l \text{var}(W^{(l)})^2 \right\}^2 + \left\{ \sum_l dCov^2(W^{(l)}) \right\}^4 \cdot \frac{1}{n^2 S^4} \rightarrow 0.$$

In view of Corollary 3.1 of [Hall & Heyde \(1980\)](#), the conclusion follows from Proposition [A.2](#) and [A.3](#). \square

Theorem [3.3](#) and Theorem [4.1](#) can be proved using similar arguments in Proposition [A.2](#) and Proposition [A.3](#), we omit the details.

A.2.3 Proof of Theorem 3.2

Proof. Under the null of mutual independence, we have

$$\begin{aligned}\mathbb{E}\hat{S}^2 &= \sum_{1 \leq l < m \leq p} \mathbb{E}[dCov_n^2(W^{(l)})dCov_n^2(W^{(m)})] \\ &= \sum_{1 \leq l < m \leq p} dCov^2(W^{(l)})dCov^2(W^{(m)}) = S^2.\end{aligned}$$

Thus it suffice to show that

$$\mathbb{E} \left(\left| \frac{\hat{S}^2}{S^2} - 1 \right|^2 \right) = \frac{\text{var}(\hat{S}^2)}{S^4} \rightarrow 0.$$

Notice that

$$\begin{aligned}\text{var}(\hat{S}^2) &= \sum_{1 \leq l < m \leq p} \sum_{1 \leq l' < m' \leq p} \text{cov}(dCov_n^2(W^{(l)})dCov_n^2(W^{(m)}), dCov_n^2(W^{(l')})dCov_n^2(W^{(m')})) \\ &= \sum_{l < m} \text{var}(dCov_n^2(W^{(l)})dCov_n^2(W^{(m)})) \\ &\quad + 2 \sum_{l < m < m'} \text{cov}(dCov_n^2(W^{(l)})dCov_n^2(W^{(m)}), dCov_n^2(W^{(l)})dCov_n^2(W^{(m')})) \\ &= \sum_{l < m} \text{var}(dCov_n^2(W^{(l)}))\text{var}(dCov_n^2(W^{(m)})) \\ &\quad + \sum_{l \neq m} \text{var}(dCov_n^2(W^{(l)}))dCov^4(W^{(m)}) \\ &\quad + 2 \sum_{l < m < m'} \text{var}(dCov_n^2(W^{(l)}))dCov^2(W^{(m)})dCov^2(W^{(m')}) \\ &:= J_1 + J_2 + J_3 \quad (\text{say}).\end{aligned}$$

Since $dCov_n^2(W^{(l)})$ is a fourth order U-statistics, by the Hoeffding decomposition, the dominant term of its variance is

$$\binom{n}{4}^{-1} \binom{4}{1} \binom{n-4}{3} \text{var}(h_1(W^{(l)}))$$

with

$$\begin{aligned}\text{var}(h_1(W^{(l)})) &= \frac{1}{4} \text{var}(\mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 | W^{(l)}]) \\ &= \frac{1}{4} \{ \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_l(W^{(l)}, W''^{(l)})^2] - dCov^4(W^{(l)}) \}.\end{aligned}$$

Under Assumption A1 and by Lemma A.1, we can derive that

$$\frac{\sum_{l=1}^p \mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_l(W^{(l)}, W''^{(l)})^2]}{nS^2} \lesssim \frac{\sum_l \{ \mathbb{E}[|W^{(l)} - \mu^{(l)}|]^2 \text{var}(W^{(l)}) \}}{nS^2} \rightarrow 0,$$

and

$$\frac{\sum_{l=1}^p dCov^4(W^{(l)})}{S^2} \rightarrow 0,$$

as we have shown in the proof of Theorem 3.1. The higher order terms of the variance of $dCov_n^2(W^{(l)})$ can be handled in a similar fashion. Hence we have

$$\frac{\sum_{l=1}^p \text{var}(dCov_n^2(W^{(l)}))}{S^2} = O\left(\frac{\sum_{l=1}^p \{\mathbb{E}[U_l(W^{(l)}, W'^{(l)})^2 U_l(W^{(l)}, W''^{(l)})^2] - dCov^4(W^{(l)})\}}{nS^2}\right) \rightarrow 0.$$

Therefore, we obtain that

$$\frac{J_1}{S^4} \leq \left[\frac{\sum_{l=1}^p \text{var}(dCov_n^2(W^{(l)}))}{S^2} \right]^2 \rightarrow 0,$$

and

$$\frac{J_2}{S^4} \leq \left(\frac{\sum_{l=1}^p \text{var}(dCov_n^2(W^{(l)}))}{S^2} \right) \cdot \left(\frac{\sum_{l=1}^p dCov^4(W^{(l)})}{S^2} \right) \rightarrow 0,$$

and also

$$\frac{J_3}{S^4} \leq \frac{2S^2 \sum_{l=1}^p \text{var}(dCov_n^2(W^{(l)}))}{S^4} \rightarrow 0.$$

Thus \hat{S}^2 is ratio consistent under the null and Assumption A1. \square

A.2.4 Proof of Theorem 3.4

Proof. When $W^{(l)}$ is standard Gaussian, we can directly calculate that $dCov^2(W^{(l)}) = f(1) = \frac{4}{\pi}(1 + \frac{\pi}{3} - \sqrt{3})$. Therefore,

$$S^2 = \sum_{1 \leq l < m \leq p} dCov^2(W^{(l)})dCov^2(W^{(m)}) = \frac{p(p-1)}{2}[f(1)]^2.$$

Our test $\phi_{n,\alpha} = 1$ if $D_n > z_\alpha$, where z_α is $100(1 - \alpha)\%$ quantile of standard normal. Hence we have

$$\begin{aligned} 1 - \mathbb{E}[\phi_{n,\alpha}] &= P\left(\frac{\sum_{1 \leq l < m \leq p} \sqrt{\binom{n}{2}} dCov_n^2(W^{(l)}, W^{(m)})}{S} \leq z_\alpha\right) \\ &= P\left(\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)}) - |\Theta|^2 \leq z_\alpha f(1) \sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2\right) \\ &\leq P\left(\left|\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)}) - |\Theta|^2\right| \geq \left|z_\alpha f(1) \sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2\right|\right) \\ &\leq \frac{\text{var}[\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)})]}{\left(z_\alpha f(1) \sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2\right)^2}, \end{aligned}$$

where $z_\alpha f(1)\sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2$ is negative for large enough \tilde{c} and the last inequality uses the fact that $\mathbb{E}[\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)})] = |\Theta|^2$ and Chebyshev's inequality. Now let $Z_i^{(lm)} = (W_i^{(l)}, W_i^{(m)})$, by lemma 2.1 we have

$$\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)}) := \frac{1}{\binom{n}{4}} \sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq n} h^s(W_{i_1}, W_{i_2}, W_{i_3}, W_{i_4}),$$

where

$$h^s(W_{i_1}, W_{i_2}, W_{i_3}, W_{i_4}) = \sum_{1 \leq l < m \leq p} h(Z_{i_1}^{(lm)}, Z_{i_2}^{(lm)}, Z_{i_3}^{(lm)}, Z_{i_4}^{(lm)}).$$

Therefore $\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)})$ is a fourth order U-statistic with kernel h^s and its variance is given by

$$\text{var} \left[\sum_{1 \leq l < m \leq p} dCov_n^2(W^{(l)}, W^{(m)}) \right] = \binom{n}{4}^{-1} \sum_{c=1}^4 \binom{4}{c} \binom{n-4}{4-c} \text{var}(h_c^s) \leq C \sum_{c=1}^4 \text{var}(h_c^s) n^{-c},$$

for some constant C . Here $h_c^s = \sum_{1 \leq l < m \leq p} h_c(z_1^{(lm)}, \dots, z_c^{(lm)})$ for $c = 1, 2, 3, 4$ with $h_c(z_1^{(lm)}, \dots, z_c^{(lm)}) = Eh(z_1^{(lm)}, \dots, z_c^{(lm)}, Z_{c+1}^{(lm)}, \dots, Z_4^{(lm)})$ defined in Appendix A.1

Use the results from Lemma A.2 or similar arguments from the proof, we can work out the variance of the fourth order U-statistic. Specifically, for some constant c' , we have

$$\begin{aligned} 4 \cdot \text{var}(h_1^s) &= \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[h_1(Z_{i_1}^{(lm)}, Z_{i_2}^{(lm)}, Z_{i_3}^{(lm)}, Z_{i_4}^{(lm)}) h_1(Z_{i_1}^{(l'm')}, Z_{i_5}^{(l'm')}, Z_{i_6}^{(l'm')}, Z_{i_7}^{(l'm')})] - |\Theta|^4 \\ &= \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)}) U(W^{(m)}, W'^{(m)}) U(W^{(l')}, W''^{(l')}) U(W^{(m')}, W''^{(m')})] - |\Theta|^4 \\ &\leq c' \left\{ p^4 \mathbb{E}[U(W^{(1)}, W'^{(1)}) U(W^{(2)}, W'^{(2)}) U(W^{(3)}, W''^{(3)}) U(W^{(4)}, W''^{(4)})] \right. \\ &\quad + p^3 \mathbb{E}[U(W^{(1)}, W'^{(1)}) U(W^{(1)}, W''^{(1)}) U(W^{(2)}, W'^{(2)}) U(W^{(3)}, W''^{(3)})] \\ &\quad + p^2 \mathbb{E}[U(W^{(1)}, W'^{(1)}) U(W^{(1)}, W''^{(1)}) U(W^{(2)}, W'^{(2)}) U(W^{(2)}, W''^{(2)})] \left. \right\} \\ &= O(|\Theta|^4) + O(|\Theta|^3) + O(|\Theta|^2). \end{aligned}$$

Therefore, we have

$$\frac{Cn^{-1} \text{var}(h_1^s)}{\left(z_\alpha f(1) \sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2 \right)^2} \leq \frac{Cn^{-1} \{O(|\Theta|^4) + O(|\Theta|^3) + O(|\Theta|^2)\}}{z_\alpha^2 f(1)^2 \frac{p(p-1)}{n(n-1)} + |\Theta|^4 - 2z_\alpha f(1) \sqrt{\frac{p(p-1)}{n(n-1)}} |\Theta|^2}.$$

Hence, the right hand side can be made less than $\frac{1-\beta}{4}$ when $p/n \rightarrow \lambda \in (0, \infty)$, and also

$|\Theta|^2 > \tilde{c}$ for some large enough constant $\tilde{c} = \tilde{c}(\alpha, \beta, \lambda)$. Similarly, we have

$$\begin{aligned} \text{var}(h_2^s) \leq & c' \left\{ \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)})U(W^{(m)}, W'^{(m)})U(W^{(l')}, W'^{(l')})U(W^{(m')}, W'^{(m')})] \right. \\ & + \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)})U(W^{(m)}, W'^{(m)})U(W^{(l')}, W''^{(l')})U(W^{(m')}, W''^{(m')})] \\ & \left. + \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)})U(W^{(m)}, W''^{(m)})U(W'''^{(l')}, W'^{(l')})U(W'''^{(m')}, W''^{(m')})] \right\}. \end{aligned}$$

In particular

$$\begin{aligned} & \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)})U(W^{(m)}, W'^{(m)})U(W^{(l')}, W'^{(l')})U(W^{(m')}, W'^{(m')})] \\ \leq & c' \left\{ p^4 \mathbb{E}[U(W^{(1)}, W'^{(1)})U(W^{(2)}, W'^{(2)})U(W^{(3)}, W'^{(3)})U(W^{(4)}, W'^{(4)})] \right. \\ & + p^3 \mathbb{E}[U(W^{(1)}, W'^{(1)})^2 U(W^{(2)}, W'^{(2)})U(W^{(3)}, W'^{(3)})] \\ & \left. + p^2 \mathbb{E}[U(W^{(1)}, W'^{(1)})^2 U(W^{(2)}, W'^{(2)})^2] \right\} \\ = & O(|\Theta|^4) + O(p|\Theta|^2) + O(p^2), \end{aligned}$$

and also

$$\begin{aligned} & \sum_{\substack{1 \leq l < m \leq p \\ 1 \leq l' < m' \leq p}} \mathbb{E}[U(W^{(l)}, W'^{(l)})U(W^{(m)}, W''^{(m)})U(W'''^{(l')}, W'^{(l')})U(W'''^{(m')}, W''^{(m')})] \\ \leq & c' \left\{ p^4 \mathbb{E}[U(W^{(1)}, W'^{(1)})U(W^{(2)}, W''^{(2)})U(W'''^{(3)}, W'^{(3)})U(W'''^{(4)}, W''^{(4)})] \right. \\ & + p^3 \mathbb{E}[U(W^{(1)}, W'^{(1)})U(W'''^{(1)}, W''^{(1)})U(W^{(2)}, W''^{(2)})U(W'''^{(3)}, W'^{(3)})] \\ & + p^3 \mathbb{E}[U(W^{(1)}, W'^{(1)})U(W'''^{(1)}, W'^{(1)})U(W^{(2)}, W''^{(2)})U(W'''^{(3)}, W''^{(3)})] \\ & \left. + p^2 \mathbb{E}[U(W^{(1)}, W'^{(1)})U(W'''^{(1)}, W'^{(1)})U(W^{(2)}, W''^{(2)})U(W'''^{(2)}, W''^{(2)})] \right\} \\ = & O(|\Theta|^4) + O(|\Theta|^3) + O(|\Theta|^2). \end{aligned}$$

Therefore,

$$\frac{Cn^{-2}\text{var}(h_2^s)}{\left(z_\alpha f(1)\sqrt{\frac{p(p-1)}{n(n-1)}} - |\Theta|^2\right)^2} \leq \frac{Cn^{-2}\{O(|\Theta|^4) + O(|\Theta|^3) + O(|\Theta|^2) + O(p|\Theta|^2) + O(p^2)\}}{z_\alpha^2 f(1)^2 \frac{p(p-1)}{n(n-1)} + |\Theta|^4 - 2z_\alpha f(1)\sqrt{\frac{p(p-1)}{n(n-1)}}|\Theta|^2}.$$

The right hand side can also be made less than $\frac{1-\beta}{4}$ when $p/n \rightarrow \lambda \in (0, \infty)$ and \tilde{c} is large.

Using similar arguments, we can show that $\text{var}(h_3^s), \text{var}(h_4^s) = O(\text{var}(h_2^s))$ and accordingly we obtain that $1 - \mathbb{E}[\phi_{n,\alpha}] \leq 1 - \beta$ as $p/n \rightarrow \lambda$ and the theorem is proved. \square

Lemma A.2. *For multivariate Gaussian (W_1, W_2, W_3, W_4) with pairwise correlation ρ , we have*

$$\begin{aligned}
\mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W'_3)U(W_4, W'_4)] &\leq C'|\rho|^4, \\
\mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W''_3)U(W_4, W''_4)] &\leq C'|\rho|^4, \\
\mathbb{E}[U(W_1, W'_1)U(W_1, W''_1)U(W_2, W'_2)U(W_3, W'_3)] &\leq C'|\rho|^3, \\
\mathbb{E}[U(W_1, W'_1)U(W_1, W''_1)U(W_2, W'_2)U(W_2, W''_2)] &\leq C'|\rho|^2, \\
\mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3''', W'_3)U(W_4''', W''_4)] &\leq C'|\rho|^4, \\
\mathbb{E}[U(W_1, W'_1)U(W_1''', W'_1)U(W_2, W'_2)U(W_3''', W''_3)] &\leq C'|\rho|^3, \\
\mathbb{E}[U(W_1, W'_1)U(W_1''', W'_1)U(W_2, W'_2)U(W_2''', W''_2)] &\leq C'|\rho|^2, \\
\mathbb{E}[U(W_1, W'_1)^2U(W_2, W'_2)U(W_3, W'_3)] &\leq C'|\rho|^2,
\end{aligned} \tag{21}$$

for some positive constant C' which is different from line to line.

Proof. We provide the details for (21). The other inequalities can be obtained in a similar way. Using Lemma 1 in Szekeley et al. (2007), we can show that

$$U(W_1, W'_1) = \int_{\mathbb{R}} (f(t_1) - e^{it_1 W_1})(\overline{f(t_1)} - e^{-it_1 W_1}) \frac{dt_1}{\pi t_1^2},$$

where $\overline{f(t)} = f(t) = e^{-t^2/2}$. Therefore,

$$\begin{aligned}
&\mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W'_3)U(W_4, W'_4)] \\
&= \mathbb{E} \left\{ \int_{\mathbb{R}^4} \pi^{-4} (e^{-t_1^2/2} - e^{it_1 W_1})(e^{-t_2^2/2} - e^{it_2 W_2})(e^{-t_3^2/2} - e^{it_3 W_3})(e^{-t_4^2/2} - e^{it_4 W_4}) \right. \\
&\quad \times (e^{-t_1^2/2} - e^{it_1 W'_1})(e^{-t_2^2/2} - e^{it_2 W'_2})(e^{-t_3^2/2} - e^{it_3 W'_3})(e^{-t_4^2/2} - e^{it_4 W'_4}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \Big\} \\
&= \int_{\mathbb{R}^4} \pi^{-4} \left| \mathbb{E}(e^{-t_1^2/2} - e^{it_1 W_1})(e^{-t_2^2/2} - e^{it_2 W_2})(e^{-t_3^2/2} - e^{it_3 W_3})(e^{-t_4^2/2} - e^{it_4 W_4}) \right|^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2}.
\end{aligned}$$

It is straightforward to verify that

$$\begin{aligned}
&\mathbb{E}(e^{-t_1^2/2} - e^{it_1 W_1})(e^{-t_2^2/2} - e^{it_2 W_2})(e^{-t_3^2/2} - e^{it_3 W_3})(e^{-t_4^2/2} - e^{it_4 W_4}) \\
&= e^{-\frac{t_1^2+t_2^2+t_3^2+t_4^2}{2}} (e^{-\rho t_1 t_2} + e^{-\rho t_1 t_3} + e^{-\rho t_1 t_4} + e^{-\rho t_2 t_3} + e^{-\rho t_2 t_4} + e^{-\rho t_3 t_4} \\
&\quad - e^{-\rho t_1 t_2 - \rho t_1 t_3 - \rho t_2 t_3} - e^{-\rho t_1 t_2 - \rho t_1 t_4 - \rho t_2 t_4} - e^{-\rho t_1 t_3 - \rho t_1 t_4 - \rho t_3 t_4} \\
&\quad - e^{-\rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} + e^{-\rho t_1 t_2 - \rho t_1 t_3 - \rho t_1 t_4 - \rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} - 3) \\
&= e^{-\frac{t_1^2+t_2^2+t_3^2+t_4^2}{2}} (3\rho^2 t_1 t_2 t_3 t_4 + \text{Remainder terms}),
\end{aligned}$$

where the last step uses the Taylor expansion $e^x = 1 + x + x^2/2 + \sum_3^\infty x^k/k!$. Therefore we

have

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W'_3)U(W_4, W'_4)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} \left| e^{-\frac{t_1^2+t_2^2+t_3^2+t_4^2}{2}} (3\rho^2 t_1 t_2 t_3 t_4 + \text{Remainder terms}) \right|^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} 9\pi^{-4} \rho^4 e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} dt_1 dt_2 dt_3 dt_4 \tag{22}
\end{aligned}$$

$$+ \int_{\mathbb{R}^4} 6\pi^{-4} \rho^2 t_1 t_2 t_3 t_4 e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \tag{23}$$

$$+ \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\text{Remainder terms})^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2}. \tag{24}$$

We first consider term (23). Denote $a_1 = t_1 t_2$, $a_2 = t_1 t_3$, $a_3 = t_1 t_4$, $a_4 = t_2 t_3$, $a_5 = t_2 t_4$ and $a_6 = t_3 t_4$. By the Vitali convergence theorem, we can show that

$$\begin{aligned}
& \int_{\mathbb{R}^4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\text{Remainder terms}) \frac{dt_1}{t_1} \frac{dt_2}{t_2} \frac{dt_3}{t_3} \frac{dt_4}{t_4} \\
&= \sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} \int_{\mathbb{R}^4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} \left\{ (a_1)^k + (a_2)^k + (a_3)^k + (a_4)^k + (a_5)^k + (a_6)^k \right. \\
&\quad - (a_1 + a_2 + a_4)^k - (a_1 + a_3 + a_5)^k - (a_2 + a_3 + a_6)^k - (a_4 + a_5 + a_6)^k \\
&\quad \left. + (a_1 + a_2 + a_3 + a_4 + a_5 + a_6)^k \right\} \frac{dt_1}{t_1} \frac{dt_2}{t_2} \frac{dt_3}{t_3} \frac{dt_4}{t_4} \\
&:= \sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} I_k.
\end{aligned}$$

Using the multinomial expansion, we have

$$\begin{aligned}
a(k) &:= \left\{ (a_1)^k + (a_2)^k + (a_3)^k + (a_4)^k + (a_5)^k + (a_6)^k - (a_1 + a_2 + a_4)^k - (a_1 + a_3 + a_5)^k \right. \\
&\quad \left. - (a_2 + a_3 + a_6)^k - (a_4 + a_5 + a_6)^k + (a_1 + a_2 + a_3 + a_4 + a_5 + a_6)^k \right\} \\
&= \sum_{*} \frac{k!}{k_1! k_2! k_3! k_4! k_5! k_6!} t_1^{k_1+k_2+k_3} t_2^{k_1+k_4+k_5} t_3^{k_2+k_4+k_6} t_4^{k_3+k_5+k_6},
\end{aligned}$$

where \sum_{*} denotes the summation over all $(k_1, k_2, k_3, k_4, k_5, k_6)$ such that $\sum_{i=1}^6 k_i = k$, $k_1 + k_2 + k_3 \geq 1$, $k_1 + k_4 + k_5 \geq 1$, $k_2 + k_4 + k_6 \geq 1$ and $k_3 + k_5 + k_6 \geq 1$. Since $\int_{\mathbb{R}} e^{-t^2} t^{2k+1} dt = 0$ and $0 < \int_{\mathbb{R}} e^{-t^2} t^{2k} dt < \infty$, we have $I_k > 0$ for $k \geq 3$. We first consider the case $-1/3 \leq \rho < 0$. By Hölder's inequality, we have

$$\sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} I_k = \sum_{k=3}^{\infty} \frac{|\rho|^k}{k!} I_k \leq \mathbb{E}[U(W^{(1)}, W'^{(1)})^4] < \infty,$$

which implies that

$$\sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} I_k = \sum_{k=3}^{\infty} \frac{|\rho|^k}{k!} I_k \leq \frac{|\rho|^3}{(1/3)^3} \sum_{k=3}^{\infty} \frac{(1/3)^k}{k!} I_k \leq C|\rho|^3.$$

For $0 \leq \rho \leq 1$, $\sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} I_k = \rho^3 \sum_{k=3}^{\infty} \frac{(-1)^k}{k!} I_k \rho^{k-3}$. First notice that the above power series is convergent at $\rho = 1$, that is,

$$\sum_{k=3}^{\infty} \frac{(-1)^k}{k!} I_k \rho^{k-3} = \sum_{k=3}^{\infty} \frac{(-1)^k}{k!} I_k \leq \mathbb{E}[U(W^{(1)}, W'^{(1)})^4] < \infty.$$

By the Abel theorem, the power series is continuous as a function of ρ for $\rho \in [0, 1]$ and therefore bounded. Equivalently, we can use the Abel's uniform convergence test to show the power series is uniformly convergent for all $\rho \in [0, 1]$. Hence, $\sum_{k=3}^{\infty} \frac{(-\rho)^k}{k!} I_k \leq C|\rho|^3$ for some constant C that is independent of ρ and accordingly term (23) $\leq C|\rho|^5$. Similarly, we can show that

$$\begin{aligned} & \int_{\mathbb{R}^4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\text{Remainder terms})^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\ &= \rho^6 \int_{\mathbb{R}^4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\rho^{-3} \times \text{Remainder terms})^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\ &= \rho^6 \int_{\mathbb{R}^4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} \sum_{k=6}^{\infty} (-\rho)^{k-6} J_k \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\ &\leq C|\rho|^6, \end{aligned}$$

where $J_k = \frac{1}{k_1!k_2!} a(k_1)a(k_2)$ such that $k_1 + k_2 = k$. Therefore

$$\mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W'_3)U(W_4, W'_4)] \leq C|\rho|^4.$$

Using similarly arguments, we can show that

$$\begin{aligned} & \mathbb{E}[U(W_1, W'_1)U(W_2, W'_2)U(W_3, W''_3)U(W_4, W''_4)] \\ &= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-\rho t_1 t_2} + e^{-\rho t_1 t_3} + e^{-\rho t_1 t_4} + e^{-\rho t_2 t_3} + e^{-\rho t_2 t_4} + e^{-\rho t_3 t_4} \\ & \quad - e^{-\rho t_1 t_2 - \rho t_1 t_3 - \rho t_2 t_3} - e^{-\rho t_1 t_2 - \rho t_1 t_4 - \rho t_2 t_4} - e^{-\rho t_1 t_3 - \rho t_1 t_4 - \rho t_3 t_4} - e^{-\rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} \\ & \quad + e^{-\rho t_1 t_2 - \rho t_1 t_3 - \rho t_1 t_4 - \rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} - 3)(e^{-\rho t_1 t_2} - 1)(e^{-\rho t_3 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\ &= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (3\rho^4 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\ &\leq C'|\rho|^4, \end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W_1, W''_1)U(W_2, W'_2)U(W_3, W''_3)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-t_1 t_2} + e^{-\rho t_1 t_3} + e^{-\rho t_1 t_4} + e^{-\rho t_2 t_3} + e^{-\rho t_2 t_4} + e^{-\rho t_3 t_4} \\
&\quad - e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_2 t_3} - e^{-t_1 t_2 - \rho t_1 t_4 - \rho t_2 t_4} - e^{-\rho t_1 t_3 - \rho t_1 t_4 - \rho t_3 t_4} - e^{-\rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} \\
&\quad + e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_1 t_4 - \rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} - 3)(e^{-\rho t_1 t_3} - 1)(e^{-\rho t_2 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (2\rho^3 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^3,
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W_1, W''_1)U(W_2, W'_2)U(W_2, W''_2)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-t_1 t_2} + e^{-\rho t_1 t_3} + e^{-\rho t_1 t_4} + e^{-\rho t_2 t_3} + e^{-\rho t_2 t_4} + e^{-t_3 t_4} \\
&\quad - e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_2 t_3} - e^{-t_1 t_2 - \rho t_1 t_4 - \rho t_2 t_4} - e^{-\rho t_1 t_3 - \rho t_1 t_4 - t_3 t_4} - e^{-\rho t_2 t_3 - \rho t_2 t_4 - t_3 t_4} \\
&\quad + e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_1 t_4 - \rho t_2 t_3 - \rho t_2 t_4 - t_3 t_4} - 3)(e^{-\rho t_1 t_3} - 1)(e^{-\rho t_2 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (2\rho^2 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^2,
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W_2, W''_2)U(W_3''', W'_3)U(W_4''', W''_4)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-\rho t_1 t_2} - 1)(e^{-\rho t_1 t_3} - 1)(e^{-\rho t_2 t_4} - 1)(e^{-\rho t_3 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\rho^4 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^4,
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W_1''', W'_1)U(W_2, W''_2)U(W_3''', W''_3)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-t_1 t_2} - 1)(e^{-\rho t_1 t_3} - 1)(e^{-\rho t_2 t_4} - 1)(e^{-\rho t_3 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\rho^3 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^3,
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)U(W''_1, W'_1)U(W_2, W''_2)U(W''_2, W'_2)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-t_1 t_2} - 1)(e^{-\rho t_1 t_3} - 1)(e^{-\rho t_2 t_4} - 1)(e^{-t_3 t_4} - 1) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (\rho^2 t_1^2 t_2^2 t_3^2 t_4^2 + \text{Remainder terms}) \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^2,
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}[U(W_1, W'_1)^2 U(W_2, W'_2) U(W_3, W'_3)] \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (e^{-t_1 t_2} + e^{-\rho t_1 t_3} + e^{-\rho t_1 t_4} + e^{-\rho t_2 t_3} + e^{-\rho t_2 t_4} + e^{-\rho t_3 t_4} \\
&\quad - e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_2 t_3} - e^{-t_1 t_2 - \rho t_1 t_4 - \rho t_2 t_4} - e^{-\rho t_1 t_3 - \rho t_1 t_4 - \rho t_3 t_4} - e^{-\rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} \\
&\quad + e^{-t_1 t_2 - \rho t_1 t_3 - \rho t_1 t_4 - \rho t_2 t_3 - \rho t_2 t_4 - \rho t_3 t_4} - 3)^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&= \int_{\mathbb{R}^4} \pi^{-4} e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} (2\rho t_1 t_2 t_3 t_4 + \text{Remainder terms})^2 \frac{dt_1}{t_1^2} \frac{dt_2}{t_2^2} \frac{dt_3}{t_3^2} \frac{dt_4}{t_4^2} \\
&\leq C' |\rho|^2.
\end{aligned}$$

□

References

- Adam, B.-L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z. & Wright, G. L. J. (2002), ‘Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men’, *Cancer research* **62**(13), 3609–3614.
- Anderson, T. W. (1958), *An introduction to multivariate statistical analysis*, Vol. 2, Wiley New York.
- Bickel, P. J. & Levina, E. (2008), ‘Regularized estimation of large covariance matrices’, *The Annals of Statistics* pp. 199–227.
- Cai, T. T. & Jiang, T. (2011), ‘Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices’, *The Annals of Statistics* **39**(3), 1496–1525.

- Cai, T. T. & Ma, Z. (2013), ‘Optimal hypothesis testing for high-dimensional covariance matrices’, *Bernoulli* **19**, 2359–2388.
- Chen, S. X., Zhang, L.-X. & Zhong, P.-S. (2010), ‘Tests for high-dimensional covariance matrices’, *Journal of the American Statistical Association* **105**(490).
- Comon, P. (1994), ‘Independent component analysis, a new concept?’, *Signal Process.* **36**(3), 287–314.
URL: [http://dx.doi.org/10.1016/0165-1684\(94\)90029-9](http://dx.doi.org/10.1016/0165-1684(94)90029-9)
- Fisher, T. J. (2012), ‘On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size’, *Journal of Statistical Planning and Inference* **142**(1), 312–326.
- Fisher, T. J., Sun, X. & Gallagher, C. M. (2010), ‘A new test for sphericity of the covariance matrix for high dimensional data’, *Journal of Multivariate Analysis* **101**(10), 2554–2570.
- Hall, P. (1984), ‘Central limit theorem for integrated square error of multivariate nonparametric density estimators’, *Journal of multivariate analysis* **14**(1), 1–16.
- Hall, P. & Heyde, C. C. (1980), *Martingale limit theory and its application*, Academic press.
- Han, F. & Liu, H. (2014), ‘Distribution-free tests of independence with applications to testing more structures’, *arXiv preprint arXiv:1410.4179*.
- Heller, R., Heller, Y. & Gorfine, M. (2013), ‘A consistent multivariate test of association based on ranks of distances’, *Biometrika* **100**, 503–510.
- Heller, R., Heller, Y., Kaufman, S., Brill, B. & Gorfine, M. (2016), ‘Consistent distribution free k -sample and independence tests for univariate random variables’, *Journal of Machine Learning Research* **17**, 1–54.
- Jiang, T. (2004), ‘The asymptotic distributions of the largest entries of sample correlation matrices’, *Annals of Applied Probability* pp. 865–880.
- John, S. (1971), ‘Some optimal multivariate tests’, *Biometrika* **58**(1), 123–127.
- Ledoit, O. & Wolf, M. (2002), ‘Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size’, *Annals of Statistics* pp. 1081–1102.

- Lehmann, E. L. (1999), *Elements of large-sample theory*, Springer Science & Business Media.
- Leung, D. & Drton, M. (2015), ‘Testing mutual independence in high dimensions with sums of squares of rank correlations’, *arXiv preprint arXiv:1501.01732*.
- Levina, E., Rothman, A., Zhu, J. et al. (2008), ‘Sparse estimation of large covariance matrices via a nested lasso penalty’, *The Annals of Applied Statistics* **2**(1), 245–263.
- Park, T., Shao, X., Yao, S. et al. (2015), ‘Partial martingale difference correlation’, *Electronic Journal of Statistics* **9**, 1492–1517.
- Pfister, N., Bühlmann, P., Schölkopf, B. & Peters, J. (2016), ‘Kernel-based tests for joint independence’.
URL: <http://arxiv.org/abs/1603.00285>
- Qiu, Y. & Chen, S. X. (2012), ‘Test for bandedness of high-dimensional covariance matrices and bandwidth estimation’, *The Annals of Statistics* **40**(3), 1285–1314.
- Schott, J. R. (2005), ‘Testing for complete independence in high dimensions’, *Biometrika* **92**(4), 951–956.
- Sejdinovic, D., Sriperumbudur, B. K., Gretton, A. & Fukumizu, K. (2013), ‘Equivalence of distance-based and rkhs-based statistics in hypothesis testing’, *Annals of Statistics* **41**, 2263–2291.
- Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Vol. 162, John Wiley & Sons.
- Shao, Q.-M. & Zhou, W.-X. (2014), ‘Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices’, *The Annals of Probability* **42**(2), 623–648.
- Srivastava, M. S. (2005), ‘Some tests concerning the covariance matrix in high dimensional data’, *Journal of the Japan Statistical Society* **35**(2), 251–272.
- Srivastava, M. S. (2006), ‘Some tests criteria for the covariance matrix with fewer observations than the dimension’, *Acta Comment. Univ. Tartu. Math* **10**, 77–93.
- Sun, Y. (1998), ‘The almost equivalence of pairwise and mutual independence and the duality with exchangeability’, *Probability Theory and Related Fields* **112**(3), 425–456.

- Székely, G. J. & Rizzo, M. L. (2009), ‘Brownian distance covariance’, *The annals of applied statistics* **3**(4), 1236–1265.
- Székely, G. J. & Rizzo, M. L. (2013*a*), ‘The distance correlation t -test of independence in high dimension’, *Journal of Multivariate Analysis* **117**(3), 193–213.
- Székely, G. J. & Rizzo, M. L. (2013*b*), ‘Energy statistics: A class of statistics based on distances’, *Journal of Statistical Planning and Inference* **143**(3), 1249–1272.
- Székely, G. J. & Rizzo, M. L. (2014), ‘Partial distance correlation with methods for dissimilarities’, *The Annals of Statistics* **42**(6), 2382–2412.
- Székely, G. J., Rizzo, M. L., Bakirov, N. K. et al. (2007), ‘Measuring and testing dependence by correlation of distances’, *The Annals of Statistics* **35**(6), 2769–2794.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005), ‘Sparsity and smoothness via the fused lasso’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(1), 91–108.
- Wagaman, A. & Levina, E. (2009), ‘Discovering sparse covariance structures with the isomap’, *Journal of Computational and Graphical Statistics* **18**(3), 551–572.
- Wu, W. B. & Pourahmadi, M. (2003), ‘Nonparametric estimation of large covariance matrices of longitudinal data’, *Biometrika* **90**(4), 831–844.

Table 1: Size of the tests from Example 5.1

	n	p	dCov	SC	LD $_{\tau}$	LD $_{\rho}$	CJ	HL $_{\tau}$	HL $_{\rho}$
i)	60	50	0.054	0.047	0.061	0.050	0.016	0.030	0.017
	60	100	0.054	0.049	0.062	0.051	0.010	0.026	0.010
	60	200	0.056	0.056	0.065	0.056	0.004	0.021	0.007
	60	400	0.050	0.046	0.060	0.048	0.001	0.018	0.002
	60	800	0.048	0.040	0.052	0.045	0.000	0.011	0.001
	100	50	0.054	0.049	0.054	0.052	0.021	0.033	0.021
	100	100	0.054	0.054	0.057	0.050	0.018	0.032	0.021
	100	200	0.053	0.050	0.054	0.049	0.013	0.030	0.018
	100	400	0.055	0.047	0.054	0.051	0.012	0.027	0.013
	100	800	0.057	0.052	0.060	0.056	0.005	0.020	0.008
ii)	60	50	0.057	0.053	0.062	0.054	0.028	0.027	0.016
	60	100	0.052	0.050	0.060	0.050	0.024	0.025	0.010
	60	200	0.057	0.055	0.060	0.049	0.017	0.021	0.006
	60	400	0.058	0.054	0.062	0.050	0.014	0.019	0.005
	60	800	0.061	0.055	0.057	0.050	0.009	0.014	0.002
	100	50	0.051	0.050	0.056	0.049	0.033	0.031	0.024
	100	100	0.048	0.046	0.052	0.046	0.031	0.033	0.022
	100	200	0.046	0.045	0.050	0.044	0.032	0.032	0.020
	100	400	0.048	0.049	0.050	0.046	0.022	0.026	0.013
	100	800	0.052	0.051	0.053	0.049	0.019	0.028	0.008
iii)	60	50	0.058	0.146	0.067	0.057	0.974	0.027	0.016
	60	100	0.052	0.148	0.062	0.054	1.000	0.027	0.010
	60	200	0.052	0.150	0.057	0.047	1.000	0.022	0.007
	60	400	0.057	0.147	0.060	0.049	1.000	0.021	0.004
	60	800	0.059	0.148	0.065	0.056	1.000	0.016	0.002
	100	50	0.054	0.143	0.057	0.052	0.975	0.036	0.027
	100	100	0.056	0.154	0.056	0.050	1.000	0.034	0.023
	100	200	0.055	0.160	0.054	0.048	1.000	0.031	0.016
	100	400	0.052	0.145	0.054	0.049	1.000	0.025	0.010
	100	800	0.053	0.142	0.060	0.056	1.000	0.020	0.010
iv)	60	50	0.055	0.073	0.066	0.056	0.377	0.033	0.017
	60	100	0.057	0.075	0.062	0.052	0.628	0.022	0.011
	60	200	0.058	0.067	0.061	0.053	0.888	0.026	0.009
	60	400	0.057	0.073	0.063	0.053	0.992	0.020	0.004
	60	800	0.057	0.071	0.058	0.048	1.000	0.014	0.003
	100	50	0.059	0.076	0.060	0.057	0.483	0.037	0.029
	100	100	0.060	0.075	0.059	0.052	0.774	0.034	0.020
	100	200	0.047	0.066	0.051	0.046	0.978	0.033	0.018
	100	400	0.054	0.070	0.051	0.045	1.000	0.030	0.013
	100	800	0.054	0.067	0.057	0.050	1.000	0.023	0.010

Table 2: Power of the tests from Example 5.2

case	n	p	dCov	SC	LD $_{\tau}$	LD $_{\rho}$	CJ	HL $_{\tau}$	HL $_{\rho}$
AR(1)	60	50	0.886	0.957	0.939	0.925	0.271	0.318	0.223
	60	100	0.906	0.969	0.949	0.939	0.158	0.240	0.137
	60	200	0.909	0.973	0.955	0.944	0.081	0.177	0.070
	60	400	0.909	0.973	0.957	0.949	0.029	0.105	0.031
	60	800	0.908	0.972	0.955	0.947	0.010	0.070	0.012
	100	50	0.998	1.000	1.000	1.000	0.849	0.827	0.764
	100	100	0.999	1.000	1.000	1.000	0.795	0.790	0.698
	100	200	1.000	1.000	1.000	1.000	0.705	0.727	0.594
	100	400	0.999	1.000	1.000	1.000	0.579	0.653	0.477
	100	800	0.999	1.000	1.000	1.000	0.428	0.573	0.353
Band	60	50	1.000	1.000	1.000	1.000	0.427	0.494	0.368
	60	100	0.999	1.000	1.000	1.000	0.285	0.406	0.247
	60	200	1.000	1.000	1.000	1.000	0.156	0.303	0.132
	60	400	1.000	1.000	1.000	1.000	0.065	0.196	0.056
	60	800	1.000	1.000	1.000	1.000	0.024	0.133	0.026
	100	50	1.000	1.000	1.000	1.000	0.965	0.957	0.928
	100	100	1.000	1.000	1.000	1.000	0.946	0.943	0.894
	100	200	1.000	1.000	1.000	1.000	0.905	0.927	0.831
	100	400	1.000	1.000	1.000	1.000	0.811	0.883	0.729
	100	800	1.000	1.000	1.000	1.000	0.668	0.807	0.578
Block	60	50	0.999	1.000	1.000	1.000	0.442	0.503	0.372
	60	100	1.000	1.000	1.000	1.000	0.282	0.400	0.239
	60	200	1.000	1.000	1.000	1.000	0.149	0.303	0.128
	60	400	1.000	1.000	1.000	1.000	0.065	0.191	0.058
	60	800	1.000	1.000	1.000	1.000	0.020	0.127	0.022
	100	50	1.000	1.000	1.000	1.000	0.959	0.952	0.918
	100	100	1.000	1.000	1.000	1.000	0.936	0.935	0.880
	100	200	1.000	1.000	1.000	1.000	0.899	0.919	0.830
	100	400	1.000	1.000	1.000	1.000	0.812	0.883	0.733
	100	800	1.000	1.000	1.000	1.000	0.666	0.805	0.571

Table 3: Power performance for detecting non-monotone dependence

	n	p	dCov	SC	LD $_{\tau}$	LD $_{\rho}$	CJ	HL $_{\tau}$	HL $_{\rho}$
Example 5.3	60	50	1.000	0.037	0.127	0.055	0.022	0.261	0.044
	60	100	1.000	0.038	0.121	0.057	0.014	0.299	0.032
	60	200	1.000	0.039	0.126	0.059	0.009	0.332	0.022
	60	400	1.000	0.033	0.117	0.054	0.006	0.369	0.017
	60	800	1.000	0.033	0.114	0.057	0.004	0.403	0.011
	100	50	1.000	0.036	0.123	0.049	0.032	0.285	0.059
	100	100	1.000	0.037	0.116	0.055	0.028	0.337	0.054
	100	200	1.000	0.036	0.117	0.056	0.025	0.390	0.046
	100	400	1.000	0.035	0.114	0.051	0.016	0.424	0.033
	100	800	1.000	0.037	0.115	0.054	0.013	0.464	0.025
Example 5.4	60	50	1.000	0.054	0.257	0.109	0.035	0.302	0.050
	60	100	1.000	0.054	0.266	0.109	0.030	0.336	0.033
	60	200	1.000	0.052	0.260	0.111	0.039	0.378	0.028
	60	400	1.000	0.059	0.271	0.112	0.031	0.420	0.016
	60	800	1.000	0.055	0.261	0.104	0.032	0.476	0.011
	100	50	1.000	0.049	0.264	0.109	0.046	0.334	0.062
	100	100	1.000	0.057	0.259	0.114	0.046	0.384	0.059
	100	200	1.000	0.048	0.253	0.106	0.061	0.436	0.048
	100	400	1.000	0.051	0.257	0.104	0.066	0.486	0.038
	100	800	1.000	0.052	0.252	0.107	0.083	0.535	0.030
Example 5.5	60	50	0.694	0.609	0.607	0.591	0.020	0.201	0.028
	60	100	0.769	0.728	0.718	0.706	0.015	0.200	0.018
	60	200	0.828	0.807	0.808	0.797	0.013	0.203	0.012
	60	400	0.887	0.873	0.874	0.867	0.008	0.193	0.008
	60	800	0.919	0.904	0.896	0.898	0.004	0.183	0.003
	100	50	0.771	0.609	0.617	0.593	0.027	0.390	0.067
	100	100	0.800	0.732	0.725	0.716	0.023	0.411	0.053
	100	200	0.843	0.808	0.805	0.800	0.021	0.450	0.042
	100	400	0.887	0.857	0.859	0.857	0.015	0.484	0.028
	100	800	0.918	0.902	0.901	0.898	0.011	0.502	0.020

Table 4: Size for the banded dependence tests from Example 5.6

case	n	p	$h = 5$			$h = 10$		
			dCov	CJ	HL_τ	dCov	CJ	HL_τ
i)	60	50	0.063	0.012	0.041	0.069	0.009	0.034
	60	100	0.067	0.008	0.048	0.067	0.007	0.044
	60	200	0.061	0.004	0.044	0.062	0.004	0.041
	60	400	0.060	0.002	0.043	0.061	0.002	0.042
	60	800	0.064	0.001	0.034	0.064	0.001	0.033
	100	50	0.056	0.021	0.040	0.053	0.016	0.033
	100	100	0.056	0.018	0.044	0.058	0.016	0.040
	100	200	0.055	0.015	0.042	0.057	0.014	0.042
	100	400	0.057	0.013	0.043	0.057	0.012	0.042
	100	800	0.059	0.005	0.041	0.061	0.005	0.040
ii)	60	50	0.065	0.938	0.038	0.063	0.893	0.031
	60	100	0.070	1.000	0.044	0.066	0.999	0.039
	60	200	0.058	1.000	0.042	0.062	1.000	0.040
	60	400	0.054	1.000	0.037	0.055	1.000	0.036
	60	800	0.066	1.000	0.036	0.064	1.000	0.036
	100	50	0.059	0.943	0.040	0.060	0.901	0.032
	100	100	0.053	1.000	0.048	0.058	1.000	0.043
	100	200	0.052	1.000	0.042	0.052	1.000	0.040
	100	400	0.057	1.000	0.043	0.060	1.000	0.042
	100	800	0.058	1.000	0.039	0.059	1.000	0.038
iii)	60	50	0.062	0.027	0.046	0.060	0.022	0.036
	60	100	0.064	0.020	0.044	0.062	0.017	0.040
	60	200	0.069	0.016	0.041	0.073	0.015	0.039
	60	400	0.060	0.013	0.042	0.061	0.013	0.041
	60	800	0.066	0.008	0.038	0.065	0.008	0.038
	100	50	0.060	0.029	0.042	0.058	0.024	0.034
	100	100	0.058	0.029	0.044	0.056	0.025	0.040
	100	200	0.053	0.026	0.046	0.054	0.025	0.043
	100	400	0.055	0.021	0.043	0.054	0.021	0.042
	100	800	0.056	0.017	0.040	0.057	0.017	0.040

Table 5: Power for the banded dependence tests from Example 5.7

case	n	p	$h = 5$			$h = 10$		
			dCov	CJ	HL_τ	dCov	CJ	HL_τ
i)	60	50	0.983	0.070	0.185	0.916	0.050	0.130
	60	100	0.998	0.038	0.149	0.957	0.028	0.114
	60	200	0.999	0.020	0.116	0.973	0.016	0.095
	60	400	1.000	0.005	0.072	0.980	0.004	0.063
	60	800	1.000	0.002	0.064	0.983	0.002	0.057
	100	50	1.000	0.249	0.345	0.998	0.171	0.244
	100	100	1.000	0.170	0.302	1.000	0.122	0.223
	100	200	1.000	0.101	0.215	1.000	0.073	0.159
	100	400	1.000	0.050	0.150	1.000	0.036	0.119
	100	800	1.000	0.026	0.114	1.000	0.020	0.095
ii)	60	50	0.938	0.091	0.192	0.804	0.062	0.129
	60	100	0.980	0.062	0.151	0.877	0.045	0.116
	60	200	0.993	0.037	0.115	0.902	0.028	0.094
	60	400	0.996	0.020	0.073	0.915	0.017	0.061
	60	800	0.998	0.012	0.051	0.921	0.011	0.044
	100	50	0.999	0.209	0.357	0.988	0.147	0.250
	100	100	1.000	0.163	0.285	0.997	0.123	0.212
	100	200	1.000	0.107	0.210	1.000	0.080	0.157
	100	400	1.000	0.067	0.154	1.000	0.051	0.116
	100	800	1.000	0.038	0.106	1.000	0.033	0.086
iii)	60	50	1.000	0.019	0.311	1.000	0.015	0.305
	60	100	1.000	0.014	0.359	1.000	0.013	0.356
	60	200	1.000	0.010	0.407	1.000	0.010	0.406
	60	400	1.000	0.008	0.449	1.000	0.007	0.449
	60	800	1.000	0.003	0.511	1.000	0.003	0.511
	100	50	1.000	0.023	0.310	1.000	0.018	0.304
	100	100	1.000	0.023	0.368	1.000	0.020	0.365
	100	200	1.000	0.023	0.423	1.000	0.021	0.420
	100	400	1.000	0.015	0.490	1.000	0.014	0.490
	100	800	1.000	0.010	0.537	1.000	0.009	0.536

Table 6: Comparison between proposed dCov test and dHSIC test in Section 5.3

Example	case	n	p	dCov	dHSIC
Example 5.1	(i)	60	18	0.051	0.050
		100	36	0.051	0.050
		200	72	0.055	0.000
	(ii)	60	18	0.055	0.052
		100	36	0.051	0.045
		200	72	0.049	0.000
	(iii)	60	18	0.056	0.049
		100	36	0.048	0.044
		200	72	0.049	0.000
	(iv)	60	18	0.052	0.048
		100	36	0.049	0.051
		200	72	0.052	0.000
Example 5.2	AR	60	18	0.782	0.070
		100	36	0.995	0.036
		200	72	1.000	0.000
	Band	60	18	0.990	0.091
		100	36	1.000	0.045
		200	72	1.000	0.000
Example 5.4-5.5		60	18	1.000	0.267
		100	36	1.000	0.173
		200	72	1.000	0.000
		60	18	1.000	1.000
		100	36	1.000	1.000
		200	72	1.000	1.000
Example 5.8		60	18	0.051	0.664
		100	36	0.048	0.282
		200	72	0.057	0.072